BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00796 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/CA00/00777

(22) International Filing Date: 28 June 2000 (28.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/140,820 28 June 1999 (28.06.1999) US

(71) Applicant *(for all designated States except US)*: NATIONAL RESEARCH COUNCIL OF CANADA [CA/CA]; Intellectual Property Services, Building M-58, Room EG-12, Montreal Road, Ottawa, Ontario K1A 0R6 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LOGAN, Susan, M. [CA/CA]; 62 Kilbarry Crescent, Ottawa, Ontario K1K 0H1 (CA). WAKARCHUK, Warren [CA/CA]; 11-837 Eastvale Drive, Gloucester, Ontario K1J 7T5 (CA). CONLAN, Wayne [CA/CA]; 1152 Bordeau Grove, Orleans, Ontario K1C 2M7 (CA). MONTEIRO, Mario, A. [CA/CA]; 171 O'Connor Street, Apt. 105, Ottawa, Ontario K1P 2T4 (CA). ALTMAN, Eleonora [CA/CA]; 689 Laverendrye Drive, Gloucester, Ontario K1J 7X5 (CA). HIRATSUKA, Koji [CA/CA]; 11 Hedgerow Lane, Stittsville, Ontario K2S 1C9 (CA).

(74) Agent: ANDERSON, J., Wayne; National Research Council of Canada, Intellectual Property Services, Building M-58, Room EG12, Montreal Road, Ottawa, Ontario K1A 0R6 (CA).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00796 A2

(54) Title: GLYCOSYLTRANSFERASES OF *HELICOBACTER PYLORI* AS A NEW TARGET IN PREVENTION AND TREATMENT OF *H. PYLORI* INFECTIONS

(57) Abstract: Novel isolated polynucleotides encoding glycosyltransferases involved in the biosynthesis of the lipopolysaccharide of *Helicobacter pylori*, together with recombinant DNA constructs and vectors containing polynucleotide sequences encoding such glycosyltransferases are disclosed. These nucleic acid constructs and vectors may be used for the preparation of glycosyltransferases they encode, by expressing the coding polynucleotide sequences in suitable host cells. Also disclosed are isolated polypeptides having enzymatic activity of helicobacterial glycosyltransferases. Such polypeptides are particularly useful for screening of modulators of their enzymatic activity, in particular enzymatic inhibitors having potential antibacterial activity.

GLYCOSYLTRANSFERASES OF *HELICOBACTER PYLORI* AS A NEW TARGET IN PREVENTION AND TREATMENT OF *H. PYLORI* INFECTIONS

FIELD OF THE INVENTION

The invention relates to newly identified and isolated polynucleotides and polypeptides of bacterial origin, in particular to novel polynucleotides and polypeptides related to glycosyltransferases involved in biosynthesis of lipopolysaccharides of *Helicobacter pylori*.

BACKGROUND OF THE INVENTION

Helicobacter pylori is a spiral, microaerophilic, Gram-negative bacterium infecting about 50% of the global human population, and is now recognised as the most common bacterial pathogen of humans worldwide. It is the causative agent of chronic active gastritis in all who harbour it, is responsible for the development of most gastro-duodenal ulcers, and is formally recognised as the carcinogen for certain gastric cancers (Blaser, *Gastroenterology* 102: 720-727 (1992); Parsonnet *et al*, *N. Engl. J. Med.* 325: 1127-1131 (1991)). *H. pylori* is a highly motile organism and migrates through the superficial mucus layer of the gastric lumen to colonize the underlying gastric pits and associated epithelium. The precise mechanisms by which *H. pylori* injures the gastric mucosa to elicit the aforementioned pathogenic states remains unknown, but it is clear that urease production (Eaton *et al*, *Infect. Immun.* 59: 2470-2475 (1991)) and motility are required for gastric colonisation of experimental animals. However, the development of gastro-duodenal disease clearly requires additional bacterial virulence factors (Phadnis *et al*, *Infect. Immun.* 62:1557-1565 (1994); Tummuru *et al*, *Mol. Microbiol.* 18: 867-876 (1995)). Although several bacterial adhesins and putative receptors on host epithelium have been described (Evans *et al*, *J. Bacteriol.* 175: 674-683 (1993); Boren *et al*, *Science* 262: 1892-1895 (1993);

Odenbreit *et al*, *Gut* 37 (Suppl. 1): A1 (1995)), their role in gastric colonization by *H. pylori* has not be n clearly established.

Gram-negative bacteria, such as *H. pylori*, have their bacterial cell wall covered with an outer membraneous layer consisting of lipids, proteins and lipopolysaccharides (LPS). LPS contain lipid A, a disaccharide of two phosphorylated glucosamine (GlcN) residues with attached fatty acids, and a polysaccharide attached to one of the glucosamine residues through a glycosidic bond. The polysaccharide is composed of a core of approximately 10 sugar residues followed by a repeating series of units of 3 to 5 sugars called the O side chain (O-chain). The number of repeating units in the O-chain varies from about 10 to 40. The sugars found in the O-chain vary among bacterial species, whereas the composition of the core polysaccharide is relatively constant. Lipopolysaccharides are released from bacteria undergoing lysis and are toxic to animals and humans. They are often referred to as endotoxins.

While much attention has focused on the role of bacterial and host proteins in *H. pylori* infection and immunity, the role of LPS in these processes has received less consideration (Moran, *Aliment. Pharmacol. Ther.* 10 (suppl): 39-50 (1996); Yokota *et al*, *Infect. Immun.* 66: 3006-3011 (1998); Wang *et al, Mol. Microbiol.* 31: 1265-1274 (1999)). As a major cell surface component, this molecule is well situated to selectively interact with surface components of the host. In particular, LPS could facilitate initial gastric colonisation, be responsible for biological interactions which modify the inflammatory response, and promote a chronic infection.

Comprehensive, detailed structural analysis of *H. pylori* LPS has revealed some unique features of the molecule which may account for certain aspects of *H. pylori*-induced pathogenesis (Aspinall *et al, Biochemistry* 35: 2489-2497; 2498-2504 (1996); Aspinall *et al, Eur. J. Biochem.* 248: 592-601 (1997); Monteiro *et al, J. Biol. Chem.* 273: 11533-11543 (1998)). In addition, *H. pylori* LPS, unlike typical LPS, has low endotoxic properties. Fresh clinical isolates usually display typical smooth type LPS (S-type). The O-chain polysaccharide structure of *H.*

pylori type strain (NCTC11637) LPS is composed of a type 2 *N*-acetyllactosamine (LacNAc) chain of various lengths and this O-chain may be partially α-L-fucosylated or less commonly α-D-glucosylated or α-D-galactosylated and may be terminated at the nonreducing end by Lewis blood group epitopes which mimic human cell surface glycoconjugates and glycolipids. However, it remains to be formally established if the O-chain of *H. pylori* LPS contributes to pathogenesis or generates protective immunity. For instance, the Lewis antigens present on the O-chain polysaccharide might reduce the immunogenicity of this molecule during infection, or might trigger autoimmunity. The ability to produce structurally defined truncated LPS molecules should help elucidate the biological role of LPS in *H. pylori* infection and immunity and possibly open a new approach to the treatment and prevention of *H. pylori* infections.

Known methods of prevention and treatment of *H. pylori* infections are either immunogenic or drug-based. The immunogenic approach is mostly intended to provide an immunogenic protection against the bacterium by vaccinating the individual with a usually bacterium-derived immunogen, to elicit an immune response of the organism to future *H. pylori* infections. Among many others, immunogens (antigens) derived from the LPS of *H. pylori* are known in this group of treatments (see, for example, WO 97/14782 and WO 87/07148).

According to the second approach, *H. pylori* infections are treated with antibacterial drugs or combinations of such drugs, intended to eradicate the bacterial population in the infected individual. In this group of treatments, the currently most common are so called triple therapies, in which patients are administered simultaneously two different antibiotics and an acid secretion inhibiting drug. The efficacy of these therapies varies and is often adversely affected by the developing resistance to broad spectrum antibiotics used for this purpose and by side effects of antibiotic therapies, which frequently result in termination of the therapy before completely healing the infection.

In view of the above-indicated deficiencies of the current antibiotic therapies, attempts are made to develop more specific drugs against *H. pylori*, such as

drugs modulating the activity of enzymes specific to the bacteria (see, for example, US 5,801,013 and US 5,942,409). An ideal anti-helicobacterial drug should be selective, meaning that the drug should inhibit *H. pylori* but not the bacterial population of the microflora of the lower intestine. This means that the molecular target of the drug should be unique to *H. pylori* and/or should be related to its unique phenotypic characteristics, particularly those facilitating the colonization of bacterium's natural ecological niche (the human stomach). While improving the understanding of *H. pylori* pathogenesis, the present invention provides means for developing new anti-helicobacterial drugs possessing such desirable characteristics.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides isolated and/or recombinant nucleic acids which encode certain glycosyltransferases of *Helicobacter* origin. The invention also provides recombinant DNA constructs and vectors containing polynucleotide sequences encoding such glycosyltransferases or portions thereof. These nucleic acids and constructs may be used to produce recombinant glycosyltransferases of *Helicobacter* origin by expressing the polynucleotide sequences in suitable host cells.

In another aspect, the invention provides isolated polypeptides having the enzymatic activity of glycosyltransferases of *Helicobacter* origin. Such polypeptides are useful, among other things, for the identification of modulators, in particular inhibitors of their enzymatic activity, which inhibitors are potential antimicrobial agents. Using the isolated polypeptides of the present invention, potential inhibitors of these enzymes can be screened for antimicrobial or antibiotic effects, without culturing pathogenic strains of *Helicobacter* bacteria, such as *H. pylori*.

According to one embodiment of the invention, preferred glycosyltransferases of *Helicobacter* origin are glycosyltransferases of *H. pylori* involved in the biosynthesis of the bacterial lipopolysaccharide (LPS), in particular of LPS core or LPS O-chain. Disrupting genes of such glycosyltransferases in several strains of

H. pylori resulted in mutants unable to complete the structural assembly of LPS and having as a result a reduced ability to colonize the murine stomach.

According to yet another aspect, the present invention provides novel antigens and vaccines used in immunization against *Helicobacter* bacteria, in particular *H. pylori*. The novel antigens are derived from bacteria having deactivated gene of a glycosyltransferase involved in the biosynthesis of the bacterial lipopolysaccharide, in particular of LPS core or LPS O-chain. Purified or partially purified LPS isolated from such mutants is a preferred antigen.

Other advantages, objects and features of the present invention will be readily apparent to those skilled in the art from the following detailed description of preferred embodiments in conjunction with the accompanying drawings and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows amino acid sequence alignment of glycosyltransferases from *H. pylori*, *H. influenzae*, *H. somnus* and *N. meningitidis*. Multiple sequence alignment was performed using the Clustal Alignment Programme (Higgins *et al, Gene* 73: 237-244 (1988)). Designations on the left side refer to the origin of the sequences; HP0826 of genebank AE000594 (Tomb *et al, Nature* 388:539-547 (1997)), *Haemophilus influenzae* lex2B, U05670 (Cope *et al, Mol. Microbiol.* 5: 1113-1124 (1994)), *Haemophilus somnus* lob1, U94833 (Inzana *et al, Infect. Immun.* 65: 4675-4681 (1997)) and *Neisseria meningitidis* lgtB, AAC44085 (Jennings *et al, Mol. Microbiol.* 18: 729-740 (1995). Numbers on the right side indicate amino acid positions. Gaps introduced to maximise the alignment are indicated by dashes. Shadings were obtained using the Genedoc Programme (www.cris.com/~ketchup/genedoc.shtml). Black indicates 100% identity, dark grey indicates 80% identity, and light grey indicates 60% identity.

Fig. 2 shows a complete FAB-MS spectrum of the methylated intact LPS of 26695::HP0826kan strain.

Fig. 3 is a schematic showing the chemical structure of LPS from parent strains 26695 and SS1 and isogenic mutants of HP0826, HP0159 and HP0479.

Fig. 4 shows results of CZE-MS/MS analysis (+ion mode) of delipidated LPS from *H. pylori* 26695::0159 mutant. Tandem mass spectrum of precursor ions at m/z 902 (doubly protonated ions). Separation conditions: 10 mM ammonium acetate containing 5% methanol, pH 9.0, +25 kV. For MS/MS experiments, nitrogen as a collision gas, E_{lab}: 70 eV (laboratory frame of reference).

Fig. 5 shows results of CZE-MS/MS (+ion mode) analysis of delipidated LPS from *H. pylori* 0479 mutants. Tandem mass spectrum of precursor ions at m/z 1612. Separation conditions: 10 mM ammonium acetate containing 5% methanol, pH 9.0, +25 kV. For MS/MS experiments, nitrogen as a collision gas, E_{lab}: 60 eV (laboratory frame of reference).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As used herein, the terms "identity" and "similarity" mean the degree of sequence relatedness between two or more polynucleotide or polypeptide sequences as determined by the match between strings of such sequences. "Identity" or "similarity" can be readily quantified by algorithms well known to those skilled in the art, implemented in a number of publicly available computer software packages, for example BLAST software package available from NCBI and other sources. The identity or similarity is usually expressed as a percentage of identity with respect to some reference sequence. For example, in a polynucleotide having a sequence 95% identical to a reference nucleotide sequence, 5% of the nucleotides of the reference sequence have been deleted or substituted with another nucleotide, or 5% of another nucleotides have been inserted into the reference sequence. These substitutions, insertions, and/or deletions may take place anywhere between 5' and 3' terminal positions, either interspersed individually among nucleotides of the reference sequence or in one or more contiguous groups within the reference sequence.

The term "isolated" as used herein means altered by the hand of man with respect to its natural state. For a substance occurring in nature, it means that this substance has been changed or removed from its natural environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not isolated, but the same polynucleotide or polypeptide separated from its natural matrix and coexisting materials is isolated, as the term is employed herein.

The term "polynucleotide" or "nucleic acid" refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified or modified RNA or DNA, whether single- or double-stranded. The term "polypeptide" or "protein" refers to any peptide or protein comprising at least two amino acid residues joined to each other by peptide bonds or modified peptide bonds.

The term "variant" as used herein means a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide but retains its essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. These difference are usually limited and variants of a polypeptide are closely similar overall and identical in many regions. A variant of a polynucleotide or polypeptide may be naturally occurring, such as an allelic variant, or may be prepared by mutagenesis techniques, by direct synthesis, or by other recombinant methods well known to those skilled in the art.

A "fragment" can be considered as a variant of a polynucleotide or polypeptide, having the same nucleotide or amino acid sequence as part of the reference polynucleotide or peptide. A fragment may be "free-standing" or comprised within a larger polynucleotide or polypeptide, normally as a single continuous region.

Nucleic acids referred to herein as "recombinant" are nucleic acids which have been produced by recombinant DNA methodology, including those nucleic acids that are generated by procedures which rely upon a method of artificial recombination, such as polymerase chain reaction (PCR) and/or cloning into a vector using restriction enzymes.

According to one aspect, the invention provides novel isolated polynucleotides and polypeptides, as described in greater detail below. In particular, the invention provides isolated polynucleotides and polypeptides related to glycosyltransferases involved in the biosynthesis of bacterial lipopolysaccharides of bacteria of the genus *Helicobacter*, more particularly the lipopolysaccharides of the species *Helicobacter pylori* and various strains thereof. In a preferred embodiment, the glucosyltransferases as those involved in the biosynthesis of the bacterial LPS, in particular of LPS core or LPS O-chain. Most particularly, the invention provides isolated polynucleotides and polypeptides identical over their entire lengths to sequences set out in Table 1.

Table 1. Polynucleotide and polypeptide sequences

Sequences from strain 26695 of *H. pylori*

A. polynucleotide sequence: ORF HP0826 [SEQ ID NO:1]

```
ttgcgtgttt ttgccatttc tttaaatcaa aaagtgtgcg atacatttgg tttagttttt 60
agagacacca caactttact caatagcatc aatgccaccc accaccaagc gcaaattttt 120
gatgcgattt attctaaaac ttttgaaggc gggttgcacc ccttagtgaa aaagcattta 180
caccettatt tcatcacgca aaacatcaaa gacatgggga ttacaaccaa tctcatcagt 240
gaggtttcta agttttatta cgctttaaaa taccatgcga agtttatgag cttgggggag 300
cttgggtgct atgcgagtca ttattccttg tgggaaaaat gcatagaact caatgaagcg 360
atctgtattt tagaagacga tataaccttg aaagaggatt ttaaagaggg cttggatttt 420
ttagaaaaac acatccaaga gttaggctat atccgcttga tgcatttatt gtatgatgcc 480
agtgtaaaaa gtgagccatt gagccataaa aaccacgaga tacaagagcg tgtggggatc 540
attaaagctt atagcgaagg ggtggggact caaggctatg tgatcacgcc taagattgcc 600
aaagtttttt tgaaatgcag ccgaaaatgg gttgttcctg tggatacgat aatggacgct 660
acttttatcc atggcgtgaa aaatctggtg ttacaacctt ttgtgatcgc tgatgatgag 720
caaatctcta cgatagcacg aaaagaagaa ccttatagcc ctaaaatcgc cttaatgaga 780
gaactccatt ttaaatattt gaaatattgg cagtttgtat aa                    822
```

B. polypeptide sequence deduced from sequence A [SEQ ID NO:2]

```
Leu Arg Val Phe Ala Ile Ser Leu Asn Gln Lys Val Cys Asp Thr Phe
  1               5                  10                  15
Gly Leu Val Phe Arg Asp Thr Thr Thr Leu Leu Asn Ser Ile Asn Ala
            20                  25                  30
Thr His His Gln Ala Gln Ile Phe Asp Ala Ile Tyr Ser Lys Thr Phe
        35                  40                  45
Glu Gly Gly Leu His Pro Leu Val Lys Lys His Leu His Pro Tyr Phe
    50                  55                  60
Ile Thr Gln Asn Ile Lys Asp Met Gly Ile Thr Thr Asn Leu Ile Ser
65                  70                  75                  80
Glu Val Ser Lys Phe Tyr Tyr Ala Leu Lys Tyr His Ala Lys Phe Met
                85                  90                  95
Ser Leu Gly Glu Leu Gly Cys Tyr Ala Ser His Tyr Ser Leu Trp Glu
            100                 105                 110
Lys Cys Ile Glu Leu Asn Glu Ala Ile Cys Ile Leu Glu Asp Asp Ile
        115                 120                 125
Thr Leu Lys Glu Asp Phe Lys Glu Gly Leu Asp Phe Leu Glu Lys His
    130                 135                 140
Ile Gln Glu Leu Gly Tyr Ile Arg Leu Met His Leu Leu Tyr Asp Ala
145                 150                 155                 160
Ser Val Lys Ser Glu Pro Leu Ser His Lys Asn His Glu Ile Gln Glu
                165                 170                 175
Arg Val Gly Ile Ile Lys Ala Tyr Ser Glu Gly Val Gly Thr Gln Gly
            180                 185                 190
Tyr Val Ile Thr Pro Lys Ile Ala Lys Val Phe Leu Lys Cys Ser Arg
        195                 200                 205
Lys Trp Val Val Pro Val Asp Thr Ile Met Asp Ala Thr Phe Ile His
    210                 215                 220
Gly Val Lys Asn Leu Val Leu Gln Pro Phe Val Ile Ala Asp Asp Glu
225                 230                 235                 240
Gln Ile Ser Thr Ile Ala Arg Lys Glu Glu Pro Tyr Ser Pro Lys Ile
                245                 250                 255
```

```
Ala Leu Met Arg Glu Leu His Phe Lys Tyr Leu Lys Tyr Trp Gln Phe
            260                 265                 270
Val
```

C. polynucleotide sequence: ORF HP0159 [SEQ ID NO:3]

```
atgagtatta ttattcctat tgtcatcgct tttgataatc actatgccat gccggctggc 60
gtgagcttgt attccatgct agcttgcgct aaaacagaac acccccaatc acaaaatgat 120
agtgaaaaac ttttttataa gatccactgc ctggtggata acttaagcct tgaaaaccag 180
agcaaactaa aagagactct agcccccttt agcgcttttt cgagcctaga atttttagac 240
atttcaaccc ccaatcttca cgccactcca atagaaccct ctgcgattga taaaatcaat 300
gaagcttttt tgcaactcaa tatttacgct aagactcgct tttctaaaat ggtcatgtgc 360
cgcttgtttt tggcttcttt attcccacaa tacgacaaaa tcatcatgtt tgatgcagac 420
actttgtttt taaacgatgt gagcgagagc ttttttcatcc cactagatgg ctattatttt 480
ggagcggcta aagattttgc ttccgataaa agccctaaac attttcaaat agtgcgagaa 540
aaagaccctc gtcaagcctt ttcccttttat gagcattacc ttaatgaaag cgatatgcaa 600
atcatctatg aaagcaatta taacgccggg tttttagtcg tgaatttaaa gctgtggcgt 660
gctgatcatt tagaagagcg cttactcaat ttaacccatc aaaaaggcca gtgcgtgttt 720
taccctgaac aggacctttt aacgctcgca tgctatcaaa aagttttaat cttgccttat 780
atttataaca cccacccttt catggccaat caaaaacgct tcatccctga caaaaaagaa 840
atcgtcatgc tgcattttta ttttgtagga aaaccttggg ttttacctac tttttcatat 900
tctaaagaat ggcatgagac tcttttaaaa acccctttt atgctgaata ttccgtgaaa 960
ttccttaaac aaatgacaga atgtttaagc cttaaagaca aacaaaaaac ctttgaattt 1020
cttgcccccc tactcaataa aaaaaccctt ttagaatacg tctttttttag attgaatagg 1080
attttcaaac gcttaaaaga aaaatttttt aactcttag                        1119
```

D. polypeptide sequence deduced from sequence C [SEQ ID NO:4]

```
Met Ser Ile Ile Ile Pro Ile Val Ile Ala Phe Asp Asn His Tyr Ala
 1               5                  10                  15
Met Pro Ala Gly Val Ser Leu Tyr Ser Met Leu Ala Cys Ala Lys Thr
            20                  25                  30
Glu His Pro Gln Ser Gln Asn Asp Ser Glu Lys Leu Phe Tyr Lys Ile
        35                  40                  45
His Cys Leu Val Asp Asn Leu Ser Leu Glu Asn Gln Ser Lys Leu Lys
    50                  55                  60
Glu Thr Leu Ala Pro Phe Ser Ala Phe Ser Ser Leu Glu Phe Leu Asp
65                  70                  75                  80
Ile Ser Thr Pro Asn Leu His Ala Thr Pro Ile Glu Pro Ser Ala Ile
                85                  90                  95
Asp Lys Ile Asn Glu Ala Phe Leu Gln Leu Asn Ile Tyr Ala Lys Thr
            100                 105                 110
Arg Phe Ser Lys Met Val Met Cys Arg Leu Phe Leu Ala Ser Leu Phe
        115                 120                 125
Pro Gln Tyr Asp Lys Ile Ile Met Phe Asp Ala Asp Thr Leu Phe Leu
    130                 135                 140
Asn Asp Val Ser Glu Ser Phe Phe Ile Pro Leu Asp Gly Tyr Tyr Phe
145                 150                 155                 160
Gly Ala Ala Lys Asp Phe Ala Ser Asp Lys Ser Pro Lys His Phe Gln
                165                 170                 175
Ile Val Arg Glu Lys Asp Pro Arg Gln Ala Phe Ser Leu Tyr Glu His
            180                 185                 190
Tyr Leu Asn Glu Ser Asp Met Gln Ile Ile Tyr Glu Ser Asn Tyr Asn
        195                 200                 205
Ala Gly Phe Leu Val Val Asn Leu Lys Leu Trp Arg Ala Asp His Leu
    210                 215                 220
```

```
Glu Glu Arg Leu Leu Asn Leu Thr His Gln Lys Gly Gln Cys Val Phe
225                 230                 235                 240
Tyr Pro Glu Gln Asp Leu Leu Thr Leu Ala Cys Tyr Gln Lys Val Leu
                245                 250                 255
Ile Leu Pro Tyr Ile Tyr Asn Thr His Pro Phe Met Ala Asn Gln Lys
            260                 265                 270
Arg Phe Ile Pro Asp Lys Lys Glu Ile Val Met Leu His Phe Tyr Phe
        275                 280                 285
Val Gly Lys Pro Trp Val Leu Pro Thr Phe Ser Tyr Ser Lys Glu Trp
    290                 295                 300
His Glu Thr Leu Leu Lys Thr Pro Phe Tyr Ala Glu Tyr Ser Val Lys
305                 310                 315                 320
Phe Leu Lys Gln Met Thr Glu Cys Leu Ser Leu Lys Asp Lys Gln Lys
                325                 330                 335
Thr Phe Glu Phe Leu Ala Pro Leu Leu Asn Lys Lys Thr Leu Leu Glu
            340                 345                 350
Tyr Val Phe Phe Arg Leu Asn Arg Ile Phe Lys Arg Leu Lys Glu Lys
        355                 360                 365
Phe Phe Asn Ser
    370
```

E. polynucleotide sequence: ORF HP0479 [SEQ ID NO:5]

```
atgcatgttg cttgtctttt ggctttaggg gataatctca tcacgcttag ccttttaaaa 60
gaaatcgctt tcaaacagca acaacccctt aaaatcctag gtactcgttt gactttaaaa 120
atcgccaagc ttttagaatg cgaaaaacat tttgaaatca ttcctctttt tgaaaatgtc 180
cctgcttttt atgaccttaa aaaacaaggc gtttttttgg cgatgaagga ttttttatgg 240
ttgttaaaag cgattaaaaa gcatcaaatc aaacgtttga ttttggaaaa acaggatttt 300
agaagcactt ttttagccaa attcattccc ataaccactc caaataaaga aattaaaaac 360
gtttatcaaa accgccagga gttgttttct caaatttatg ggcatgtttt tgataacccc 420
ccatatccca tgaatttaaa aaaccccaaa aagattttga tcaacccttt cacaagatcc 480
atagaccgaa gtatcccttt agagcattta caaatcgttt taaaactttt aaaacccttt 540
tgtgttacgc ttttagattt tgaagaacga tacgcttttt taaaagacag agtcgctcat 600
tatcgcgcta aaaccagttt agaagaagtt aaaaacctga ttttagaaag cgatttgtat 660
ataggagggg attcgttttt gatccatttg gcttactatt taaagaaaaa ttattttatc 720
ttttttttata gggataatga tgatttcatg ccgcctaata gtaagaataa aaattttcta 780
aaagcccaca aaagccattc tatagaacaa gatttagcca aaaaattccg ccatttgggg 840
ctattataa                                                       849
```

F. polypeptide sequence deduced from sequence E [SEQ ID NO:6]

```
Met His Val Ala Cys Leu Leu Ala Leu Gly Asp Asn Leu Ile Thr Leu
 1               5                  10                  15
Ser Leu Leu Lys Glu Ile Ala Phe Lys Gln Gln Gln Pro Leu Lys Ile
            20                  25                  30
Leu Gly Thr Arg Leu Thr Leu Lys Ile Ala Lys Leu Leu Glu Cys Glu
        35                  40                  45
Lys His Phe Glu Ile Ile Pro Leu Phe Glu Asn Val Pro Ala Phe Tyr
    50                  55                  60
Asp Leu Lys Lys Gln Gly Val Phe Leu Ala Met Lys Asp Phe Leu Trp
65                  70                  75                  80
Leu Leu Lys Ala Ile Lys Lys His Gln Ile Lys Arg Leu Ile Leu Glu
                85                  90                  95
Lys Gln Asp Phe Arg Ser Thr Phe Leu Ala Lys Phe Ile Pro Ile Thr
            100                 105                 110
Thr Pro Asn Lys Glu Ile Lys Asn Val Tyr Gln Asn Arg Gln Glu Leu
```

```
        115                 120                 125
Phe Ser Gln Ile Tyr Gly His Val Phe Asp Asn Pro Pro Tyr Pro Met
    130                 135                 140
Asn Leu Lys Asn Pro Lys Lys Ile Leu Ile Asn Pro Phe Thr Arg Ser
145                 150                 155                 160
Ile Asp Arg Ser Ile Pro Leu Glu His Leu Gln Ile Val Leu Lys Leu
                165                 170                 175
Leu Lys Pro Phe Cys Val Thr Leu Leu Asp Phe Glu Glu Arg Tyr Ala
            180                 185                 190
Phe Leu Lys Asp Arg Val Ala His Tyr Arg Ala Lys Thr Ser Leu Glu
        195                 200                 205
Glu Val Lys Asn Leu Ile Leu Glu Ser Asp Leu Tyr Ile Gly Gly Asp
    210                 215                 220
Ser Phe Leu Ile His Leu Ala Tyr Tyr Leu Lys Lys Asn Tyr Phe Ile
225                 230                 235                 240
Phe Phe Tyr Arg Asp Asn Asp Asp Phe Met Pro Pro Asn Ser Lys Asn
                245                 250                 255
Lys Asn Phe Leu Lys Ala His Lys Ser His Ser Ile Glu Gln Asp Leu
            260                 265                 270
Ala Lys Lys Phe Arg His Leu Gly Leu Leu
        275                 280
```

G. polynucleotide sequence: ORF 1191 [SEQ ID NO:7]

```
atgagcgtaa atgcacccaa acgcatgcgt attttattgc gtttgcctaa ttggttaggc 60
gatggggtga tggcaagttc gctttttac accttaaac accactaccc taacgcgcat 120
tttatcttag tgggcccaac cattacttgc gaacttttca aaaaagatga aaaaatagaa 180
gccgttttta tagacaacac caaaaaatcc tttttcaggc tgctagccat tcacaaactc 240
gctcaaaaaa tagggcgttg cgatatagcg atcactttaa acaaccattt ctattccgct 300
tttttgctct atgcgacaaa aacgcccgtt cgcatcggtt ttgctcaatt ttttcgttct 360
ttgtttctca gccatgcgat cgctcctgcc cctaaagagt atcaccaagt ggaaaagtat 420
tgctttttat tttcgcaatt tttagaaaaa gaattggatc aaaaaagcgt tttaccctta 480
aaactggcct ttaacctccc cactcacacc ccaaacaccc ctaaaaaaat cggctttaac 540
cctagcgcaa gctatgggag tgctaaaaga tggccagctt cttattacgc tgaagtttct 600
gctgttttgt tagaaaaagg gcatgaaatt tatttttttg gggctaaaga agacgctatc 660
gtttctgaag aaatttttaaa actcatcaaa ggctcattaa aaaacccctc attgttccat 720
aacgcttaca atctgtgcgg gaaaacaagc attgaagaat tgatagagcg catcgctgtt 780
ttagatttat tcatcactaa cgatagcggc cctatgcatg tggctgctag catgcaaacc 840
cccttaatcg ctcttttttgg ccccactgat gaaaaagaga ctcgccccta taaagctcaa 900
aaaacgatcg tattgaacca ccatttaagc tgtgcgcctt gcaagaaacg agtttgccct 960
ttaaagaatg caaaaaacca tttgtgcatg aaatctatca cgccccttga agtcctagaa 1020
gccgctcaca ctcttttaga agagccttaa                                  1050
```

H. polypeptide sequence deduced from sequence G [SEQ ID NO:8]

```
Met Ser Val Asn Ala Pro Lys Arg Met Arg Ile Leu Leu Arg Leu Pro
  1               5                  10                  15
Asn Trp Leu Gly Asp Gly Val Met Ala Ser Ser Leu Phe Tyr Thr Leu
             20                  25                  30
Lys His His Tyr Pro Asn Ala His Phe Ile Leu Val Gly Pro Thr Ile
         35                  40                  45
Thr Cys Glu Leu Phe Lys Lys Asp Glu Lys Ile Glu Ala Val Phe Ile
     50                  55                  60
Asp Asn Thr Lys Lys Ser Phe Phe Arg Leu Leu Ala Ile His Lys Leu
 65                  70                  75                  80
Ala Gln Lys Ile Gly Arg Cys Asp Ile Ala Ile Thr Leu Asn Asn His
                 85                  90                  95
```

```
Phe Tyr Ser Ala Phe Leu Leu Tyr Ala Thr Lys Thr Pro Val Arg Ile
          100                 105                 110
Gly Phe Ala Gln Phe Phe Arg Ser Leu Phe Leu Ser His Ala Ile Ala
      115                 120                 125
Pro Ala Pro Lys Glu Tyr His Gln Val Glu Lys Tyr Cys Phe Leu Phe
  130                 135                 140
Ser Gln Phe Leu Glu Lys Glu Leu Asp Gln Lys Ser Val Leu Pro Leu
145                 150                 155                 160
Lys Leu Ala Phe Asn Leu Pro Thr His Thr Pro Asn Thr Pro Lys Lys
                165                 170                 175
Ile Gly Phe Asn Pro Ser Ala Ser Tyr Gly Ser Ala Lys Arg Trp Pro
          180                 185                 190
Ala Ser Tyr Tyr Ala Glu Val Ser Ala Val Leu Leu Glu Lys Gly His
      195                 200                 205
Glu Ile Tyr Phe Phe Gly Ala Lys Glu Asp Ala Ile Val Ser Glu Glu
  210                 215                 220
Ile Leu Lys Leu Ile Lys Gly Ser Leu Lys Asn Pro Ser Leu Phe His
225                 230                 235                 240
Asn Ala Tyr Asn Leu Cys Gly Lys Thr Ser Ile Glu Glu Leu Ile Glu
                245                 250                 255
Arg Ile Ala Val Leu Asp Leu Phe Ile Thr Asn Asp Ser Gly Pro Met
          260                 265                 270
His Val Ala Ala Ser Met Gln Thr Pro Leu Ile Ala Leu Phe Gly Pro
      275                 280                 285
Thr Asp Glu Lys Glu Thr Arg Pro Tyr Lys Ala Gln Lys Thr Ile Val
  290                 295                 300
Leu Asn His His Leu Ser Cys Ala Pro Cys Lys Lys Arg Val Cys Pro
305                 310                 315                 320
Leu Lys Asn Ala Lys Asn His Leu Cys Met Lys Ser Ile Thr Pro Leu
                325                 330                 335
Glu Val Leu Glu Ala Ala His Thr Leu Leu Glu Glu Pro
          340                 345
```

Sequences from strain SS1 of *H. pylori*

I. polynucleotide sequence: ORF SS0826 [SEQ ID NO:9]

```
ttgcgtattt ttatcatttc tttaaatcaa aaagtgtgcg ataaatttgg tttggttttt 60
agagacacca cgactttact caatagcatc aatgccaccc accaccaagt gcaaattttt 120
gatgcgattt attctaaaac ttttgaaggc gggttgcacc ccttagtgaa aaagcattta 180
cacccttatt tcatcacgca aaacatcaaa gacatgggaa ttacaaccag tctcatcagt 240
gaggtttcta agttttatta cgctttaaaa taccatgcga agtttatgag cttgggagag 300
cttgggtgct atgcgagcca ttattccttg tgggaaaaat gcatagaact caatgaagcg 360
atctgtattt tagaagacga tataaccttg aaagaggatt ttaaagaggg cttggatttt 420
ttagaaaaac acatccaaga gttaggctat gttcgcttga tgcatttatt atatgatccc 480
aatattaaaa gtgagccatt gaaccataaa aaccacgaga tacaagagcg tgtagggatt 540
attaaagctt atagcgaagg ggtggggact caaggctatg tgatcacgcc caagattgcc 600
aaagttttta aaaaacacag ccgaaaatgg gttgttcctg tggatacgat aatggacgct 660
acttttatcc atggcgtgaa aaatctggtg ttacaacctt ttgtgatcgc tgatgatgag 720
caaatctcta cgatagcgcg aaaagaacaa ccttatagcc ctaaaatcgc cttaatgaga 780
gaactccatt ttaaatattt gaaatattgg cagtttatat ag                    822
```

J. polypeptide sequence deduced from sequence I [SEQ ID NO:10]

```
Leu Arg Ile Phe Ile Ile Ser Leu Asn Gln Lys Val Cys Asp Lys Phe
  1               5                  10                  15
```

```
Gly Leu Val Phe Arg Asp Thr Thr Thr Leu Leu Asn Ser Ile Asn Ala
            20                  25                  30
Thr His His Gln Val Gln Ile Phe Asp Ala Ile Tyr Ser Lys Thr Phe
        35                  40                  45
Glu Gly Gly Leu His Pro Leu Val Lys Lys His Leu His Pro Tyr Phe
    50                  55                  60
Ile Thr Gln Asn Ile Lys Asp Met Gly Ile Thr Thr Ser Leu Ile Ser
65                  70                  75                  80
Glu Val Ser Lys Phe Tyr Tyr Ala Leu Lys Tyr His Ala Lys Phe Met
                85                  90                  95
Ser Leu Gly Glu Leu Gly Cys Tyr Ala Ser His Tyr Ser Leu Trp Glu
            100                 105                 110
Lys Cys Ile Glu Leu Asn Glu Ala Ile Cys Ile Leu Glu Asp Asp Ile
        115                 120                 125
Thr Leu Lys Glu Asp Phe Lys Glu Gly Leu Asp Phe Leu Glu Lys His
    130                 135                 140
Ile Gln Glu Leu Gly Tyr Val Arg Leu Met His Leu Leu Tyr Asp Pro
145                 150                 155                 160
Asn Ile Lys Ser Glu Pro Leu Asn His Lys Asn His Glu Ile Gln Glu
                165                 170                 175
Arg Val Gly Ile Ile Lys Ala Tyr Ser Glu Gly Val Gly Thr Gln Gly
            180                 185                 190
Tyr Val Ile Thr Pro Lys Ile Ala Lys Val Phe Lys Lys His Ser Arg
        195                 200                 205
Lys Trp Val Val Pro Val Asp Thr Ile Met Asp Ala Thr Phe Ile His
    210                 215                 220
Gly Val Lys Asn Leu Val Leu Gln Pro Phe Val Ile Ala Asp Asp Glu
225                 230                 235                 240
Gln Ile Ser Thr Ile Ala Arg Lys Glu Gln Pro Tyr Ser Pro Lys Ile
                245                 250                 255
Ala Leu Met Arg Glu Leu His Phe Lys Tyr Leu Lys Tyr Trp Gln Phe
260                 265                 270
Ile
```

K. polynucleotide sequence: ORF SS0159 [SEQ ID NO:11]

```
atgagtatta ctattcctat tgttatcgct tttgacaatc attacgccat tccggctggc  60
gtgagcctgt attccatgct agcttgcact aaaacagaac acccccaatc acaaaatgat  120
agtgaaaaac ttttttataa aatccactgc ctggtagata acttaagcct tgaaaaccag  180
tgcaaattga aagaaactct agcccccttt agcgctttta tgagcgtgga ttttttagac  240
atttcaaccc ctaatcttta cacccccttca atagaaccct ctgcgattga taaaatcaat  300
gaagcttttt tgcaactcaa tatttacgct aagactcgct tttctaaaat ggtcatgtgc  360
cgcttgtttt tggcttcttt attcccgcaa tacgacaaaa tcatcatgtt tgatgcggac  420
actttgtttt taaacgatgt gagcgagagt ttttttatcc cgctagatgg ttattatttt  480
ggagcggcta aagatttttc ttctcctaaa aaccttaaac attttcaaac agaaagggag  540
agagagcctc gccaaaaatt ttttctccat gagcattacc ttaaagaaaa agacatgaaa  600
atcatttgtg aaaaccacta taatgttggg tttttaatcg tgaatttaaa gctgtggcgt  660
gctgatcatt tagaagaacg cttactcaat ttaacccatc aaaaaggcca gtgtgtgttt  720
tgccctgaac aggatatttt aacgctcgca tgctatcaaa aagtttaca attaccttat  780
atttacaaca cccacccttt catggtcaat caaaaacgct tcatccctaa caaaaaagaa  840
atcgtcatgc tgcattttta ttttgtagga aaaccttggg ttttacccac tgctttatat  900
tctaaagaat ggcatgagac tcttttaaaa acccctttt acgctgaata ttccgtgaaa  960
tttcttaaac aaatgacaga atttttaagc cttaaagaca aacaaaaaac ctttgaattt  1020
cttgcccccc tactcaataa aaaaaccctt ttagaatatg tctttttttag attgaatagg  1080
attttcaaac gcttaaaaga aaaactttta aactcttagc                         1120
```

L. polypeptide sequence deduced from sequence K [SEQ ID NO:12]

```
Met Ser Ile Thr Ile Pro Ile Val Ile Ala Phe Asp Asn His Tyr Ala
 1               5                  10                  15
Ile Pro Ala Gly Val Ser Leu Tyr Ser Met Leu Ala Cys Thr Lys Thr
            20                  25                  30
Glu His Pro Gln Ser Gln Asn Asp Ser Glu Lys Leu Phe Tyr Lys Ile
        35                  40                  45
His Cys Leu Val Asp Asn Leu Ser Leu Glu Asn Gln Cys Lys Leu Lys
    50                  55                  60
Glu Thr Leu Ala Pro Phe Ser Ala Phe Met Ser Val Asp Phe Leu Asp
65                  70                  75                  80
Ile Ser Thr Pro Asn Leu Tyr Thr Pro Ser Ile Glu Pro Ser Ala Ile
                85                  90                  95
Asp Lys Ile Asn Glu Ala Phe Leu Gln Leu Asn Ile Tyr Ala Lys Thr
            100                 105                 110
Arg Phe Ser Lys Met Val Met Cys Arg Leu Phe Leu Ala Ser Leu Phe
        115                 120                 125
Pro Gln Tyr Asp Lys Ile Ile Met Phe Asp Ala Asp Thr Leu Phe Leu
    130                 135                 140
Asn Asp Val Ser Glu Ser Phe Phe Ile Pro Leu Asp Gly Tyr Tyr Phe
145                 150                 155                 160
Gly Ala Ala Lys Asp Phe Ser Ser Pro Lys Asn Leu Lys His Phe Gln
                165                 170                 175
Thr Glu Arg Glu Arg Glu Pro Arg Gln Lys Phe Phe Leu His Glu His
            180                 185                 190
Tyr Leu Lys Glu Lys Asp Met Lys Ile Ile Cys Glu Asn His Tyr Asn
        195                 200                 205
Val Gly Phe Leu Ile Val Asn Leu Lys Leu Trp Arg Ala Asp His Leu
    210                 215                 220
Glu Glu Arg Leu Leu Asn Leu Thr His Gln Lys Gly Gln Cys Val Phe
225                 230                 235                 240
Cys Pro Glu Gln Asp Ile Leu Thr Leu Ala Cys Tyr Gln Lys Val Leu
                245                 250                 255
Gln Leu Pro Tyr Ile Tyr Asn Thr His Pro Phe Met Val Asn Gln Lys
            260                 265                 270
Arg Phe Ile Pro Asn Lys Lys Glu Ile Val Met Leu His Phe Tyr Phe
        275                 280                 285
Val Gly Lys Pro Trp Val Leu Pro Thr Ala Leu Tyr Ser Lys Glu Trp
    290                 295                 300
His Glu Thr Leu Leu Lys Thr Pro Phe Tyr Ala Glu Tyr Ser Val Lys
305                 310                 315                 320
Phe Leu Lys Gln Met Thr Glu Phe Leu Ser Leu Lys Asp Lys Gln Lys
                325                 330                 335
Thr Phe Glu Phe Leu Ala Pro Leu Leu Asn Lys Lys Thr Leu Leu Glu
            340                 345                 350
Tyr Val Phe Phe Arg Leu Asn Arg Ile Phe Lys Arg Leu Lys Glu Lys
        355                 360                 365
Leu Leu Asn Ser
    370
```

M. polynucleotide sequence: ORF SS0479 [SEQ ID NO:13]

```
atgcatgttg cttgtctttt ggctttaggg gataacctca tcacgcttag cctttgtgaa 60
gaaatcgctc tcaaacagca acaacccctt aaaatcctag gtactcgttt gactttaaaa 120
atcgccaagc ttttagaatg cgaaaaacat tttgaaatca ttcctgtttt taaaaatatc 180
cccgcttttt atgaccttaa aaaacaaggc gtttttggg cgatgaagga ttttttatgg 240
```

```
ttattaaaag cgcttaagaa gcacaaaatc aaacacttga ttttagaaaa acaagatttt 300
agaagcgctc ttttatccaa atttgtttcc ataaccactc caaataaaga aattaaaaat 360
gcttatcaaa accgccagga gttgttttct caaatttatg ggcatgtttt tgataatagt 420
caatattcca tgagtttaaa aaaccccaaa aagattttaa tcaacccttt cacgagagaa 480
aataatagaa atatttcttt agaacatttg caaatcgttt taaaactgtt aaaacccttt 540
tgtgttacgc ttttagattt tgaagaacga tacgcttttt taaaagatga agtcgctcat 600
tatcgcgcta aaaccagttt agaagaagct aaaaacctga ttttagaaag cgatttgtat 660
atagggggg attcgttttt gatccatttg gcttactatt taaagaaaaa ttattttatc 720
ttttttttata gggataatga cgatttcatg ccgcctaaga atgaaaattt tctaaaagcc 780
cataaaagcc atttcataga gcaggattta gccacccagt tccgccattt ggggctatta 840
taa                                                               843
```

N. polypeptide sequence deduced from sequence M [SEQ ID NO:14]

```
Met His Val Ala Cys Leu Leu Ala Leu Gly Asp Asn Leu Ile Thr Leu
  1               5                  10                  15
Ser Leu Cys Glu Glu Ile Ala Leu Lys Gln Gln Gln Pro Leu Lys Ile
             20                  25                  30
Leu Gly Thr Arg Leu Thr Leu Lys Ile Ala Lys Leu Leu Glu Cys Glu
         35                  40                  45
Lys His Phe Glu Ile Ile Pro Val Phe Lys Asn Ile Pro Ala Phe Tyr
     50                  55                  60
Asp Leu Lys Lys Gln Gly Val Phe Trp Ala Met Lys Asp Phe Leu Trp
 65                  70                  75                  80
Leu Leu Lys Ala Leu Lys Lys His Lys Ile Lys His Leu Ile Leu Glu
                 85                  90                  95
Lys Gln Asp Phe Arg Ser Ala Leu Leu Ser Lys Phe Val Ser Ile Thr
            100                 105                 110
Thr Pro Asn Lys Glu Ile Lys Asn Ala Tyr Gln Asn Arg Gln Glu Leu
        115                 120                 125
Phe Ser Gln Ile Tyr Gly His Val Phe Asp Asn Ser Gln Tyr Ser Met
    130                 135                 140
Ser Leu Lys Asn Pro Lys Lys Ile Leu Ile Asn Pro Phe Thr Arg Glu
145                 150                 155                 160
Asn Asn Arg Asn Ile Ser Leu Glu His Leu Gln Ile Val Leu Lys Leu
                165                 170                 175
Leu Lys Pro Phe Cys Val Thr Leu Leu Asp Phe Glu Glu Arg Tyr Ala
            180                 185                 190
Phe Leu Lys Asp Glu Val Ala His Tyr Arg Ala Lys Thr Ser Leu Glu
        195                 200                 205
Glu Ala Lys Asn Leu Ile Leu Glu Ser Asp Leu Tyr Ile Gly Gly Asp
    210                 215                 220
Ser Phe Leu Ile His Leu Ala Tyr Tyr Leu Lys Lys Asn Tyr Phe Ile
225                 230                 235                 240
Phe Phe Tyr Arg Asp Asn Asp Asp Phe Met Pro Pro Lys Asn Glu Asn
                245                 250                 255
Phe Leu Lys Ala His Lys Ser His Phe Ile Glu Gln Asp Leu Ala Thr
            260                 265                 270
Gln Phe Arg His Leu Gly Leu Leu
        275                 280
```

Sequences from strain PJ1 of *H. pylori*

O. polynucleotide sequence: ORF PJ1-0479 [SEQ ID NO:15]

```
atgcatgttg cttgtctttt ggctttaggg gataacctca tcacgcttag ccttttaaaa 60
gaaatcgctt ccaaacagca acggcccctt aaaatcctag gcactcgttt gactttaaaa 120
```

```
atcgccaagc ttttagaatg cgaaaaacat tttgaaatca ttcctatttt tgaaaatatc 180
cctgcttttt atgatcttaa aaaacaaggc gtttttggg cgatgaagga ttttttatgg 240
ttgttaaaag caattaagaa gcacaaaatc aaacatttga ttttagaaaa acaagatttt 300
agaagttttc ttttatccaa atttgtttcc ataaccactc ccaataaaga aattaaaaac 360
gtttatcaaa accgccagga gttgttttct ccaatttatg ggcatgtttt tgataacccc 420
ccatatccca tgaatttaaa aaaccccaaa aagattttga tcaacccttt cacaagatcc 480
atagagcgaa gtatcccttt agagcattta aaaatcgttt taaaactctt aaaacccttt 540
tgtgttacgc ttttagattt tgaagaacga tacgcttttt tacaaaatga agccactcat 600
tatcgtgcta aaaccagttt agaagaagtt aaaagcctga ttttagaaag cgatttgtat 660
atagggggg attcgttttt aatccatttg gcttactatt taaagaaaaa ttattttatc 720
ttttttttata gggataatga cgatttcatg ccacctaatg gtaagaagga aaattttcta 780
aaagcccaca aaagccatta catagaacag gatttagcca aaaaattccg ccatttgggg 840
cttattataa                                                       850
```

P. polypeptide sequence deduced from sequence O [SEQ ID NO:16]

```
Met His Val Ala Cys Leu Leu Ala Leu Gly Asp Asn Leu Ile Thr Leu
 1               5                  10                  15
Ser Leu Leu Lys Glu Ile Ala Ser Lys Gln Gln Arg Pro Leu Lys Ile
            20                  25                  30
Leu Gly Thr Arg Leu Thr Leu Lys Ile Ala Lys Leu Leu Glu Cys Glu
        35                  40                  45
Lys His Phe Glu Ile Ile Pro Ile Phe Glu Asn Ile Pro Ala Phe Tyr
    50                  55                  60
Asp Leu Lys Lys Gln Gly Val Phe Trp Ala Met Lys Asp Phe Leu Trp
65                  70                  75                  80
Leu Leu Lys Ala Ile Lys Lys His Lys Ile Lys His Leu Ile Leu Glu
                85                  90                  95
Lys Gln Asp Phe Arg Ser Phe Leu Leu Ser Lys Phe Val Ser Ile Thr
            100                 105                 110
Thr Pro Asn Lys Glu Ile Lys Asn Val Tyr Gln Asn Arg Gln Glu Leu
        115                 120                 125
Phe Ser Pro Ile Tyr Gly His Val Phe Asp Asn Pro Pro Tyr Pro Met
    130                 135                 140
Asn Leu Lys Asn Pro Lys Lys Ile Leu Ile Asn Pro Phe Thr Arg Ser
145                 150                 155                 160
Ile Glu Arg Ser Ile Pro Leu Glu His Leu Lys Ile Val Leu Lys Leu
                165                 170                 175
Leu Lys Pro Phe Cys Val Thr Leu Leu Asp Phe Glu Glu Arg Tyr Ala
            180                 185                 190
Phe Leu Gln Asn Glu Ala Thr His Tyr Arg Ala Lys Thr Ser Leu Glu
        195                 200                 205
Glu Val Lys Ser Leu Ile Leu Glu Ser Asp Leu Tyr Ile Gly Gly Asp
    210                 215                 220
Ser Phe Leu Ile His Leu Ala Tyr Tyr Leu Lys Lys Asn Tyr Phe Ile
225                 230                 235                 240
Phe Phe Tyr Arg Asp Asn Asp Asp Phe Met Pro Pro Asn Gly Lys Lys
                245                 250                 255
Glu Asn Phe Leu Lys Ala His Lys Ser His Tyr Ile Glu Gln Asp Leu
            260                 265                 270
Ala Lys Lys Phe Arg His Leu Gly Leu Ile Ile
        275                 280
```

Preferred embodiments of the invention are polynucleotides coding for *H. pylori* glycosyltransferases involved in the biosynthesis of the core or O-chain regions of the bacterial lipopolysaccharide (LPS), in particular polynucleotides having sequences shown in Table 1 (SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 and 15), polynucleotides closely related thereto, as well as fragments and variants thereof. Another preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to polynucleotides shown in Table 1, preferably at least 80% identical, more preferably at least 90% identical, most preferably at least 95% identical, and polynucleotides that are complementary to such polynucleotides. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the most preferred.

Of the polynucleotides showing substantial identity to the polynucleotides shown in Table 1 (SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 and 15), preferred are those which encode polypeptides showing substantially the same biological function or activity as the polypeptides shown in Table 1 (SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14 and 16).

Polynucleotides shown in Table 1 correspond to open reading frames HP0826 (SEQ ID NO: 1), HP0159 (SEQ ID NO: 3), HP0479 (SEQ ID NO: 5) and HP1191 (SEQ ID NO:7) of the genomic DNA of *H. pylori* strain 26695, to open reading frames SS0826 (SEQ ID NO: 9), SS0159 (SEQ ID NO: 11) and SS0479 (SEQ ID NO: 13) of the genomic DNA of *H. pylori* strain SS1, and to open reading frame PJ1-0479 (SEQ ID NO:15) of the genomic DNA of *H. pylori* strain PJ1. Among several others, ORFs HP0826, HP0159, HP0479 and HP1191 have been identified using the complete annotated genome sequence of *H. pylori* strain 26695 and BLAST analysis as potentially coding for glycosyltransferases. They have been proven, directly or indirectly, to encode a β-1,4-galactosyltransferase (HP0826), a α-1,6-glucosyltransferase (HP0159), a heptosyltransferase (HP0479), and an ADP-heptose-LPS heptosyltransferase II (HP1191), which are enzymes involved in the biosynthesis of the *H. pylori* lipopolysaccharide. ORFs

identified by BLAST analysis have been cloned, expressed, and isolated using techniques well known to those skilled in the art, also discussed more in detail further in this disclosure.

The isolated polynucleotides of the present invention can be used in the production of polypeptides they encode. For example, a polynucleotide containing all or part of the coding sequence for a *Helicobacter* glycosyltransferase can be incorporated into various DNA constructs, such as expression cassettes, and vectors, such as recombinant plasmids, adapted for further manipulation of polypeptide sequences or for the production of the encoded polypeptide in suitable host cells, either eukaryotic, such as yeast or plant cells, or prokaryotic, such as bacteria, for example *E. coli*. This can be achieved using recombinant DNA techniques and methodologies well known to those skilled in the art.

Thus the present invention further provides recombinant nucleic acids comprising polynucleotide sequences which encode glycosyltransferases involved in the biosynthesis of lipopolysaccharides of bacteria of the genus *Helicobacter*, more particularly of lipopolysaccharides of the species *Helicobacter pylori* and various strains thereof. Most particularly, the invention provides recombinant nucleic acids comprising polynucleotides identical over their entire lengths to polynucleotides having sequences set out in Table 1, as well as fragments and variants of such sequences. Among fragments and variants, preferred are those coding for polypeptides retaining the biological function or activity of the reference polypeptides.

The isolated polynucleotides and fragments thereof can also be used as DNA diagnostic probes specific to *H. pylori*, for diagnostic or similar purposes. They may be used, for example, to check whether or not the polynucleotides according to the present invention are transcribed in bacteria of an infected tissue. They may be also useful in diagnosis of the stage of infection and determining the specific pathogen involved.

The isolated polynucleotides of the present invention may further be used as hybridization probes for RNA, cDNA and genomic DNA, for example to isolate cDNA or genomic clones of other genes that have a high sequence similarity to the polynucleotides of the present invention. Such probes will comprise at least 15 bases, preferably at least 30 bases, but may have even more than 50 bases.

Preferred isolated or recombinant polypeptides of the present invention are those showing the activity of glycosyltransferases involved in biosynthesis of the bacterial lipopolysaccharides of bacteria of the genus *Helicobacter*, more particularly lipopolysaccharides of the species *Helicobacter pylori* and various strains thereof. Most particularly preferred are polypeptides coded by polynucleotides having sequences shown in Table 1 (SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13 and 15), and also those which have at least 50% identity to polypeptides shown in Table 1 (SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14 and 16), preferably at least 70% identity, more preferably at least 80% identity, most preferably at least 95% identity, polypeptides closely related thereto as well as fragments and variants thereof. Of the polypeptides having substantial identity to polypeptides of Table 1, preferred are those having the same biological function or activity as the polypeptides appearing in Table 1.

Polypeptides having amino acid sequences shown in Table 1 correspond to those coded by open reading frames HP0826 (SEQ ID NO: 2), HP0159 (SEQ ID NO: 4), HP0479 (SEQ ID NO: 6) and HP1191 (SEQ ID NO:8) of the genomic DNA of *H. pylori* strain 26695, by open reading frames SS0826 (SEQ ID NO: 10), SS0159 (SEQ ID NO: 12) and SS0479 (SEQ ID NO: 14) of the genomic DNA of *H. pylori* strain SS1, and by open reading frame PJ0479 of the genomic DNA of *H. pylori* strain PJ1. Among several others, these ORFs have been cloned and expressed in suitable host cells and their function has been determined *in vitro* using techniques well known to those skilled in the art and discussed more in detail further in this disclosure.

Polypeptides of the present invention can be produced as discussed above in connection with recombinant nucleic acids of the present invention. They can be

recovered and purified from recombinant cell cultures by methods and techniques well known to those skilled in the art, including ammonium sulfate or thanol precipitation, acid extraction, and various forms of chromatography, in particular ion exchange and high performance liquid chromatography (HPLC). Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denaturated during isolation and/or purification.

The invention also relates to methods of screening compounds, to identify those which enhance (agonists) or block (antagonists) the action of polynucleotides or polypeptides of the present invention. Of those, antagonists acting as bacteriostatic or bactericidal agents are of particular interest. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the present invention and therefore inhibit its activity. Polynucleotides and polypeptides of the present invention may be used to assess the binding of small molecule substrates and ligands from various sources, including cells, cell-free preparations, chemical libraries, and natural product mixtures. The substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics.

Polypeptides of the present invention are particularly useful for screening chemical compounds modulating the enzymatic activity of glycosyltransferases of *Helicobacter* origin involved in the biosynthesis of bacterial lipopolysaccharides, to identify those which enhance (agonists) or inhibit (antagonists or inhibitors) the action of *Helicobacter* glycosyltransferases, in particular compounds that are bacteriostatic and/or bactericidal. The method of screening may involve high-throughput techniques and assays. In a typical assay, a synthetic reaction mix comprising a polypeptide of the present invention and a labelled substrate or ligand of such polypeptide is incubated in the absence and in the presence of a candidate substance, a potential agonist or antagonist of the enzyme under study. This capability is reflected in decreased binding of the labeled ligand or in decreased production of a product from the labeled substrate. Detection of the rate or level of production of the product from the substrate may be enhanced by

using a suitable reporter system, such as a colorimetrically labelled substrate which is converted into a colorimetrically assayable product or a reporter gene responsiv to changes in the enzymatic activity of the polypeptide.

The polypeptides of the present invention showing enzymatic activity of *Helicobacter* glycosyltransferases are also useful for the enzymatic synthesis of bacterial lipopolysaccharides and fragments thereof. When included in suitable reaction mixtures, these polypeptides catalyze the transfer of mono- or oligosaccharide residues to a suitable acceptor. In a preferred embodiment, the polypeptides of the present invention are used for the preparation of various mimics, analogues and derivatives of *Helicobacter* lipopolysaccharides.

In yet another aspect, the invention provides novel mutants of *Helicobacter* bacteria, in particular mutants of *H. pylori*, having mutated (deactivated) genes of glycosyltransferases involved in the biosynthesis of bacterial lipopolysaccharides, in particular of the core or O-chain regions of LPS. Structural analysis of LPS isolated from the mutants confirmed that O-chain synthesis has been affected by the mutations and revealed the exact structure of the truncated LPS molecules. The mutant strains were also shown to have a reduced capacity of gastric colonization.

The mutant bacteria expressing the truncated LPS and the LPS isolated from such mutants are useful as sources of antigens to be used in vaccination against *Helicobacter* bacteria, in particular against *H. pylori*. Such vaccines are normally prepared from dead bacterial cells, using methods well known to those skilled in the art, and usually contain various auxiliary components, such as an appropriate adjuvant and a delivery system. A delivery system aiming at mucosal delivery is preferred. Preferably but not essentially, the antigenic preparation is administered orally to the host, but parenteral admistration is also possible. Live vaccines based on *H. pylori* mutants may also be prepared, but would normally require an appropriate vector for mucosal delivery. Vaccines of the present invention are useful in preventing and reducing the number of *H. pylori* infections and indirectly

in reducing the incidence of pathological conditions associated with such infections, in particular gastric cancer.

Chemically modified LPS isolated from mutants expressing the truncated LPS is a preferred antigen for use in vaccines according to the present invention. It is isolated from the bacteria and at least partially purified using techniques well known to those skilled in the art. Preparations of at least 70%, particularly 80%, more particularly 90%, most particularly 95% pure LPS are preferred. The purity of an LPS preparation is expressed as the weight percentage of the total Helicobacter antigens present in the preparation. The purified LPS can be used as antigen either directly or after being conjugated to a suitable carrier protein.

In the following, the invention will be described in still greater detail, by way of examples and with respect to the preferred embodiments.

Identification and cloning of β-1,4-galactosyltransferase

A search of the *H. pylori* genomic database of translated proteins revealed three open reading frames (ORFs) (HP0826, HP0805 and HP0619) which exhibited limited homology with the *lex2B* gene from *Haemophilus influenzae* (39% identity) and the *lob1* gene from *Haemophilus somnus* (32% identity). While both the *lex2B* and *lob1* genes of *Haemophilus* have been shown to be involved in synthesis of the outer core region of the lipooligosaccharide (Jarosik *et al, Infect. Immun.* 62: 4861-4867 (1994); Inzana *et al, Infect. Immun.* 65: 4675-4681 (1997)), to date no definitive function for either gene has been proposed. There is evidence that they are involved in addition of glucose (*lex2B*) and galactose (*lob1*) to the core heptose region. Both *lex2B* and *lob1* show significant homology to a larger group of LOS biosynthesis proteins which include the *H. influenzae lex1/lic2A* genes (Cope *et al, Mol. Microbiol.* 5: 1113-1124 (1994)) and *lic2B* gene (High *et al, Mol. Microbiol.* 9: 1275 (1993)), *Neisseria lgtB* and *lgtE* genes (Wakarchuk *et al, J. Bio. Chem.* 271: 19166-19173 (1996)) and *lpsA* of *P. haemolytica* (Potter *et al, FEMS Microbiol. Lett.* 129: 75-81 (1995) which are all involved in outer core assembly. The LgtB and LgtE proteins of *N. meningitidis*

have been shown to be galactosyltransferases involved in the transfer of galactose in a β-1,4 linkage in the terminal lacto-N-neotetraose structure. LgtB is responsible for the addition of Gal to GlcNAc, an identical function to that described here for HP0826, while LgtE catalyses the addition of Gal to Glc (Wakarchuk *et al*, *supra*). Clustal multiple sequence alignment of HP0826 amino acid (aa) sequence and lex2B, lob1 and lgtB aa sequences from this group of related LOS biosynthesis proteins did identify two regions of conservation spanning the regions in HP0826 from approx. aa90 to aa142 and aa189 to aa235 (see Fig 1). Limited homology was also observed with waaX from *E. coli* (Heinrichs *et al*, *Mol. Microbiol.* 30: 221-232 (1998)), a putative core β-1,4-galactosyltransferase, only in the region spanning aa96-aa142 (data not shown). No significant homology was obtained with any putative glycosyltransferases involved in O-chain assembly from Gram-negative bacteria.

Synthetic oligonucleotide primers which contained BamHI restriction sites which flanked the 5' and 3' ends of HP0826, HP0619, and HP0805 respectively, were used in a PCR reactions containing chromosomal DNA of *H. pylori* 26695 or SS1 as a template. A single PCR product was obtained in each case and this was cloned into pUC19 to give plasmids pHP0826, pHP0805, and pHP0619. DNA sequencing was used to confirm the identity of the cloned PCR products from 26695 and SS1.

Three additional open reading frames of *H. pylori* genome, HP0159, HP1191 and HP0479, have been identified by BLAST analysis as potentially coding for LPS glycosyltransferases. Of those, HP0159 displayed homology to the *rfaJ*, lipopolysaccharide 1,2-glucosyltransferase gene from a number of bacterial species, HP0479 and HP1191 displayed homology to *waaC* and *waaF* respectively, which are heptosyltransferase genes responsible for the addition of LD heptose to KDO in the core backbone.

Functional analysis of lex2B homologues

β-1,4-galactosyltransferase activity has previously been detected in *H. pylori* (Chan *et al*, *Glycobiology* 5: 683-688 (1995)), but the gene(s) for this enzyme

have not been described. Enzyme activity was detected in extracts of *E. coli* pHP0826 but not from clones of HP0805 and HP0619 using the synthetic acceptor molecule FCHASE aminophenylβ-GlcNAc and UDP-Gal as the donor. The lack of detectable activity in HP0805 and HP0619 clones could be a lack of the appropriate donor/acceptor molecule for their respective enzymatic activities. β-1,4-galactosyltransferase activity was also present in parent *H. pylori* strains but not in the *H. pylori* HP0826 mutants. The assays were followed by TLC analysis of the reaction mixtures as previously described (Gilbert *et al, Eur. J. Biochem.* 249: 187-194 (1997)). A more sensitive capillary electrophoresis (CE) analysis of the reaction mixtures clearly demonstrated a loss of galactosyltransferase activity in the mutants. The product of the enzymatic reaction had an identical CE mobility compared to a known FCHASE-aminophenyl-β-N-acetyllactosamine standard, and was subjected to NMR analysis to determine the linkage. The ^{1}H and ^{13}C chemical shift data (Table 2) and 1D NOE analysis were consistent with the linkage of the Gal being β-1,4 to the GlcNAc. The product was also sensitive to β-galactosidase.

Table 2. Linkage analysis of the product formed by HP0826 encoded protein. ^{1}H and ^{13}C chemical shifts of the glycoside of Gal-β-1,4-GlcNAc-β-FEX[a]

Residue	Position	H	C
A, β-GlcNAc	1	4.86	100.6
	2	3.91	55.8
	3	3.72	73.4
	4	3.72	79.0
	5	3.46	75.8
	6	3.74, 3.83	60.8
	NAc	1.91	22.9
B, β-Gal	1	4.46	103.8
	2	3.58	72.0
	3	3.68	73.4
	4	3.94	69.4
	5	3.73	76.3
	6	3.77, 3.77	62.0
FEXas		3.09	29.4
FEXms		2.80	36.9
FEXxs		3.57	37.6
FEXa1		6.92	118.2
FEXx1		7.28	124.4
FEXa2		7.17	132.5
EXm2		7.70	123.3
FEXx2		8.00	121.5
FEXa3		7.22	132.7
FEXa3'		7.13	131.1
FEXm3		6.82	121.5
FEXx3		6.91	104.3

[a] in ppm from the 600 MHz HSQC spectrum of the sample in D_2O at 35°C. Chemical shifts are referenced to the methyl resonance of acetone set at 2.225 ppm for ^{1}H and 31.07 ppm ^{13}C. The error is ± 0.03 ppm for ^{1}H and ± 0.3 for ^{13}C chemical shifts.
The AMX spin system of CH_2-CH_2-S-CH_2 is at 3.09, 2.80, 3.57 ppm with J_{AM}=6.4 Hz and with their respective ^{13}C signals at 29.4, 36.9 and 37.6 ppm. The aminophenyl A_2X_2 spin system is at 6.92 and 7.28 ppm with J_{AX}=8.7 Hz and their respective ^{13}C signals at 118.2 and 124.4 ppm. The three AMX spin system for fluorescein carboxamido group with J_{AM}=8-9 Hz and J_{MX}= 1-2 Hz are at (7.17, 7.70, 8.00), (7.22, 6.82, 6.91) and (7.13, 6.82, 6.91) ppm. Their respective ^{13}C signals are at (132.5, 123.3, 121.5), (132.7, 121.5, 104.3) and (131.1, 121.5, 104.3) ppm.

Functi nal analysis of rfaJ homologue (HP0159)

Enzyme activity was detected in extracts of *E. coli* pHP0159 using the synthetic acceptor molecule FCHASE aminophenyl-α-maltose or FCHASE aminophenyl-α-glucose and UDP-Glc as the donor. Activity was also present in parent *H. pylori* strains but not in *H. pylori* HP0159 mutants. The assays were followed by TLC and CE analysis of the reaction mixtures as previously described (Gilbert *et al, Eur. J. Biochem.* 249: 187-194 (1997)). The more sensitive capillary electrophoresis (CE) analysis of the reaction mixtures demonstrated a loss of glucosyltransferase activity in the mutants. The product of the enzymatic reaction was subjected to NMR analysis to determine the linkage (Table 3). The ^{1}H and ^{13}C chemical shift data, and 1D NOE analysis were consistent with the linkage of Glc being α-1,6 to the Glc.

Table 3. Linkage analysis of the product formed by HP0159 encoded protein. ^{1}H and ^{13}C chemical shifts of Glc-α-1,6-Glc-α-1,6-Glc-α-FEX[a]

Residue	Position	H	C
A, α-Glc-FEX	1	5.35	98.3
	2	3.62	72.1
	3	3.80	74.1
	4	3.48	70.6
	5	3.72	72.1
	6	3.43, 3.69	66.5
B, α-Glc	1	4.74	98.8
	2	3.47	72.2
	3	3.61	74.3
	4	3.48	70.6
	5	3.73	71.2
	6	3.59, 3.87	66.5
C, α-Glc (terminal)	1	4.89	98.8
	2	3.52	72.5
	3	3.70	74.1
	4	3.41	70.5
	5	3.67	72.8
	6	3.74,3.79	61.5
FEXas		3.02	29.3
FEXms		2.74	36.9
FEXxs		3.52	37.5
FEXa1		7.00	118.6
FEXx1		7.27	124.2
FEXa2		6.92	131.9
FEXm2		7.60	124.6
FEXx2		8.07	120.7
FEXa3		6.95	132.0
FEXa3'		6.92	131.9
FEXm3		6.69	119.6
FEXx3		6.79	104.1

[a] in ppm from the 600 MHz HSQC spectrum of the sample in D_2O at 40°C. Chemical shifts are referenced to the methyl resonance of acetone set at 2.225 ppm for ^{1}H and 31.07 ppm for ^{13}C. The error is ± 0.03 ppm for ^{1}H and ± 0.3 for ^{13}C chemical shifts.

Functional analysis of waaF homologu (HP1191)

Complementation analysis was used to determine the function of the HP1191 from *Helicobacter pylori* strain 26695. The *H. pylori* HP1191 gene was amplified by PCR (see Table 6 for primer sequences used) and cloned into pUC19 to obtain pHP1191. WaaF mutant strain *S. typhimurium* 3789 was electroporated with this recombinant plasmid, and one of the resultant transformants selected for further study. SDS-PAGE was used to analyze LPS molecules produced by the relevant *S. typhimurium* strains. The LPS of the wild type strain formed the ladder like pattern indicative of the presence of the O antigen repeat unit whereas the LPS of the *S. typhimurium* waaF mutant appeared as a single fast migrating band. The migration pattern of this mutant was not affected by the presence of the plasmid vector. However, when the *H. pylori* gene HP1191 was present *in trans* in strain 3789, this *S. typhimurium* mutant synthesized an LPS which migrated in a pattern identical to that obtained with the LPS of the wild type strain. This confirmed the activity of HP1191 to be involved in catalyzing the addition of a second heptose molecule onto the heptose linked directly to KDO in the core.

Construction of *H. pylori* mutants carrying a disrupted HP0826 gene

In order to determine the role of the HP0826 ORF in LPS biosynthesis, *H. pylori* mutants carrying a disrupted HP0826 gene were constructed by allelic exchange. Briefly, the HP0826 ORF cloned in pUC19 was disrupted by using reverse primers 5'TACAGATCGCTTCATTGAGTTCT3' and 5'CCAAGAGTTAGGCTATATCCGCTT3' in a PCR reaction and ligating a kanamycin resistance cassette (or Km^r) to the gel purified product to make plasmid pHP0826::kan. *H. pylori* strains 26695, NCTC11637, O:3 and Sydney strain (SS1) were transformed with plasmid pHP0826::kan DNA following the procedure of Haas *et al, Mol. Microbiol.* 8:753-760 (1993). This construct contains 515bp of homologous DNA upstream of the mutation and 464bp downstream of the mutation. Following transformation, cells were plated on blood agar containing kanamycin (20 µg/ml). Km^r colonies were isolated and passaged on the same medium. Individual colonies were selected and screened for the presence of a double cross over mutation in the chromosome of the kan mutant.

To assess the insertion site of the disrupted gene PCR analysis was us d, with chromosomal DNA from parent and mutant *H. pylori* strains as templates and the primer pair 5'ACACTGGCATCATACAAT3' and 5'CCATGCGAAGTTTATGAGCT3' which are internal in the structural gene. This analysis demonstrated conclusively that the Km^r cassette was inserted into the chromosomal copy of HP0826. The primer pair amplified the expected 212bp fragment in the parent strain, but resulted in a 1.6kb fragment consistent with insertion of the 1.4kb Km^r cassette. Plasmid vector sequences were not detected by Southern blotting and a single 1.7kb Hind III fragment corresponding to insertion of the kan cassette in the HP0826 ORF was present in chromosomal DNA's of 26695::0826kan mutant and SS1::0826kan mutant but not in parental DNA when probed with the kan cassette. These data confirm that the insertion mutant was the result of a double cross-over event. Four kanamycin resistant transformants were independently tested to verify that gene disruption and gene replacement had occurred. All four mutants grew normally *in vitro* (as assessed by OD vs viable numbers) and produced a truncated LPS as assessed by electrophoretic mobility on SDS-PAGE gels. The overall protein composition of the total membrane fraction was unchanged in the knockout mutants of SS1 and 26695 as assessed by SDS-PAGE and Coomassie blue staining. The contribution of polar effects to the phenotype of the HP0826 mutant is highly unlikely as a transcriptional terminator lies immediately downstream of the HP0826 ORF, the transcriptional organization switches strands and the downstream annoted ORF HP0827 is unrelated to LPS biosynthesis.

The construction of *H. pylori* mutants carrying disrupted HP0159 and HP0479 genes was carried out in essentially the same manner as above.

Genomic Organization and Allelic Variation of SS1

To ascertain if the structural organization found in 26695 and J99 is conserved within the SS1 genome, PCR amplification and sequencing of the HP0826 homologue and flanking sequence was obtained from SS1. As with 26695 and J99, the upstream and downstream ORFs are conserved although variation in the intervening sequence was evident. Allelic variation of SS1 HP0826 resulted

in 31 base pair differences between SS1 and 26695 and 46 base pair differences between SS1 and J99. These differences in DNA sequence results in a total of 9 amino acid changes in the SS1 protein when compared with 26695 and J99 amino acid sequences. In both comparisons the variations were located predominately at the N and C terminal region of the protein.

SDS-PAGE analysis of *H. pylori* HP0826 mutants

To characterize the effect of the HP0826 mutation on LPS structure in *H. pylori*, proteinase K digested whole cell lysates from both parent and mutant cells grown in broth were analyzed by SDS-PAGE. Silver staining revealed significant differences in the electrophoretic mobility of LPS isolated from parent and mutant cells of each strain examined. LPS from strains 26695, SS1, O:3 and NCTC11637 appeared to have typical high molecular weight, smooth form LPS (S-LPS), while the HP0826 mutant of each strain no longer produced the S-LPS, but appeared to produce a semi-rough type LPS. Immunoblotting with monoclonal antibodies to Lewis X (Le^x) and Lewis Y (Le^y) antigens confirmed that the LPS from all mutants no longer displayed immunoreactive material of high molecular weight typical of the corresponding parental O-chain which displays Lewis antigens.

SDS-PAGE analysis of *H. pylori* HP0159, 0479 and 1191 mutants

To characterize the effect of the HP0159, 0479 and 1191 mutations on LPS structure in *H. pylori*, proteinase K digested whole cell lysates from both parent and mutant cells grown in broth were analyzed by SDS-PAGE. Silver staining revealed significant differences in the electrophoretic mobility of LPS isolated from parent and mutant cells of each strain examined. In all cases, LPS from mutant cells no longer produced S-type LPS but instead only a fast migrating rough type LPS was observed.

Structural investigations of *H. pylori* HP0826 LPS mutants of strains 26695, SS1, and NCTC 11637

The LPS molecules of *H. pylori* strains 26695, SS1 (M. A. Monteiro *et al, Eur. J. Biochem.* 267: 305-320 (2000) and type strain NCTC 11637 (Aspinall *et al,*

supra) have been determined to carry O- chain regions that xpress Le^x and Le^y blood-group determinants. These Lewis-mimicking O chains were shown to be covalently connected to a core oligosaccharide. Sugar composition analysis (Table 4) of the intact LPSs of *H. pylori* 26695::HP0826kan, SS1::HP0826kan and NCTC 11637::HP0826kan demonstrated a clear reduction in levels of those sugars known to form the O chain components, namely L-Fuc, D-Gal and D-GlcNAc, when compared to parent LPSs.

Table 4. Approximate molar ratios of the alditol acetate derivatives of 26695, SS1 and NCTC 11637 HP0826 isogenic mutants intact LPSs. Numbers in parentheses indicate ratios obtained for respective parent strains. Analyses performed on LPS prepared from broth grown cells.

Strain	L-Fuc	D-Glc	D-Gal	GlcNAc	DD-Hep	LD-Hep
26695::Hp0826kan	0.8 (6)	6 (7)	1 (10)	1 (8)	2 (2)	1.8 (1.6)
SS1::Hp0826kan	0.8 (6)	2 (2)	1 (10)	1 (8)	2 (2)	1.8 (1.6)
NCTC11637::Hp0826kan	0.8 (6)	6 (7)	1 (10)	1 (8)	2 (2)	1.8 (1.6)

Methylation linkage analysis performed on the intact *H. pylori* mutant LPSs from each strain showed the presence of terminal and 3-substituted Fuc, terminal, 3-, and 6-(except in SS1 strain) substituted Glc, terminal, 3- and 4-substituted Gal, 2-(only in 26695), 3-(only in 26695), 6-(only in 26695), 7- and 2,7-substituted DD-Hep, 2- and 3,7-substituted LD-Hep, and terminal and 3-substituted GlcNAc units. All sugars were present in the pyranose conformation. In order to obtain sugar sequence information of the outer-extremities of the LPS molecule (O-chain perimeter), a fast atom bombardment-mass spectrometry (FAB-MS) experiment in the positive ion mode was carried out on the methylated intact mutant LPSs from each strain. The FAB-MS spectra showed several A-type primary glycosyl oxonium ions of defined composition. The trace amounts of terminal GlcNAc that were observed in the linkage analyses were also detected in each of the three mutant LPS FAB-MS spectra at m/z 260 $[GlcNAc]^+$ (Fig. 2). A-type primary glycosyl oxonium ions containing L wis blood-group related Fuc

and GlcNAc residues were observed at m/z 434 [Fuc, GlcNAc]$^+$, 508 [GlcNAc, Hep]$^+$, and 682 [Fuc, GlcNAc, Hep]$^+$. The ion m/z 434 stood for a disaccharide composed of Fuc and GlcNAc and ion m/z 508 characterized a possible connection between the O-chain related GlcNAc and a heptose from the core region. The ion m/z 682 [Fuc, GlcNAc, Hep]$^+$ represented a moiety containing GlcNAc and Fuc residues of the O-chain region and a single heptose unit from the core region which bridges the O-chain and the core oligosaccharide. Since no terminal Hep unit was detected, these m/z 508 and 682 ions must originate from cleavage at the heptose glycosidic bond and represent a partial O-chain repeating unit [Fuc, GlcNAc, Hep]$^+$. No 3,4-substituted GlcNAc, 2-substituted Gal and no m/z 638 (characteristic of Le^x) and 812 (characteristic of Le^y) glycosyl oxonium ions were detected, and therefore no evidence of an O-chain containing Le^x or Le^y determinants was found in these analyses of 26695::HP0826kan, SS1::HP0826kan and NCTC 11637::HP0826kan LPSs. In addition, higher mass ions in the FAB-MS spectrum of NCTC11637::HP0826kan at m/z 886 [Fuc, GlcNAc, Hep, Glc]$^+$, 1090 [Fuc, GlcNAc, Hep, Glc_2]$^+$, and 1294 [Fuc, GlcNAc, Hep, Glc_3]$^+$ (Fig. 2) represented the characteristic glucosylated by a [(1-6)-α-glucan] heptose unit (Aspinall *et al*, *supra*) in strain NCTC 11637 and 26695 (Fig. 2). The same primary ions were also observed in the FAB-MS spectrum of the methylated LPS of 26695::HP0826kan, but not of SS1::HP0826kan, in line with the structural findings in the parent strains (M. A. Monteiro, unpublished). In the three FAB-MS spectra, the primary ion m/z 668 and its corresponding secondary ion m/z 228 (Fig. 2) pointed to the presence of the type 1 linear B blood-group [Gal-(1-3)-Gal-(1-3)-GlcNAc] antigen, a blood-group determinant found in the LPSs of 26695, SS1 (M. A. Monteiro, unpublished), and in NCTC 11637 (Monteiro *et al, J. Biol. Chem.* 273: 11533-11543 (1998)). The glycose units emanating from the core oligosaccharide regions were of the same type as those found in the respective parent LPSs. The GlcNAc and Fuc units observed were remnants of an incomplete O chain. A comparison of the structures identified in parent and mutant LPS from 26695 and SS1 and the respective HP0826,0159 and 0479 isogenic mutants is presented in Fig 3.

Structural characterization of *H. pylori* LPS mutants 26695::HP0159kan and SS1:: HP0159kan

Growth of bacterial strains and isolation of LPS by hot aqueous phenol method were carried out as described previously (Logan *et al*, *Mol. Microbiol.* 35: 1156-1167 (2000)). Sugar analysis of the intact LPS of *H. pylori* 26695:: HP0159kan, SS1:: HP0159kan, O:3:: HP0159kan showed significant reduction in L-Fuc, D-Gal, and DD-Hep (for serotype O:3 mutant) when compared with the parent LPS indicating the presence of the structure devoid of O-chain and DD-heptan. Methylation analysis of the intact LPS from each strain showed the presence of terminal and 3-substituted L-Fuc, terminal and 4-substituted D-Glc, terminal, 3- and 4-substituted D-Gal, terminal, 2-, 6-, 7- and 2,7-substituted DD-Hep, terminal, 2- and 3-substituted LD-Hep and terminal, 3-substituted and 4-substituted D-GlcNAc. All sugars were present in the pyranose form. In addition, methylation analysis of LPS from 26695::HP0159kan and O:3::HP0159kan revealed the presence of 4-substituted D-Glc, no 6-substituted D-Glc was observed. NMR analysis of a high molecular mass fraction, isolated by gel filtration chromatography from a partially delipidated LPS (1.5% acetic acid, 1h, 100°C) from 26695:: HP0159kan by gel filtration chromatography, indicated it to contain β-1,4-linked glucan, a contaminant produced by some strains of *H. pylori* (Knirel *et al*, *Eur. J. Biochem.* 266: 123-131 (2000)). In order to deduce the sequence information on the outer extremities of the LPS molecule, permethylated intact LPS from each strain was subjected to the fast-atom-bombardment mass spectrometric analysis in the positive mode. A-type primary glycosyl oxonium ions containing Lewis blood group related Fuc and GlcNAc residues were observed at m/z 260 $[GlcNAc]^+$ and m/z 682 $[Fuc,GlcNAc, Hep]^+$. No higher mass ions representing a glucosylated DD-heptose unit were detected. This evidence together with the absence of 6-substituted glucose in methylation analysis indicated this LPS mutant to be deficient in the biosynthesis of α(1-6)–glucan present in both 26695 and O:3 parent strains. Absence of the 3-substituted glucose in methylation analysis of LPS from 26695::HP0159kan, SS::HP0159kan, suggested that addition of a 1,3-linked glucopyranosyl residue was also impaired by this mutation. In the three FAB-MS spectra, the primary ion m/z 668 and its corresponding secondary ion m/z 228 pointed to the presence of

the type 1 linear B blood group [Gal(1-3)Gal(1-3)GlcNAc] antigen, a blood group antigen found in the LPS of 26695 and SS1 (Monteiro *et al*, *Eur. J. Biochem.* 267:305-320 (2000)). Other Lewis blood group-related secondary ions were observed at m/z 228 (260-32) $[GlcNAc]^+$, 402 (434-32) $[Fuc,GlcNAc]^+$, 576 (608-32) [Fuc (1-3)Fuc (1-4)GlcNAc]$^+$ as previously described (Monteiro *et al*, *J. Biol. Chem.* 273: 11533-11543 (1998), Logan *et al*, *Mol. Microbiol.* 35: 1156-1167 (2000)).

LPS from 26695::HP0159kan was treated with 0.1 M sodium acetate buffer, pH 4.2 (2 h, 100^{o}C) and following the removal of lipid A by low speed centrifugation, subjected to the gel filtration chromatography on a Bio-Gel P-2 column, followed by capillary electrophoresis-electrospray mass spectrometry (CE-ESMS) as described previously (Thibault and Richards, *Meth. Mol. Biol.* 145: 327-343 (2000)). The CE-ESMS spectrum of the delipidated LPS confirmed the presence of a major glycoform produced by the 26695::HP0159 mutant LPS, corresponding to $FucGlcNAcHex_2Hep_4$(PE)KDO (m/z 902, doubly protonated ion). MS-MS of the doubly charged ion (m/z 902) (Fig. 4) afforded a singly charged fragment at m/z 1601 consistent with the loss of GlcNAc (and its anhydro form at m/z 1583) which subsequently lost Fuc and Hep residues to afford a fragment ion at m/z 1262. A comparison of the structures identified in parent and HP0159 mutant LPS is presented in Fig. 3.

Structural characterization of *H. pylori* LPS mutants 26695::HP0479kan and SS1::HP0479kan.

Sugar analysis of the HP0479 LPS mutants indicated reduction in the amount of L-Fuc, D-Gal and DD-Hep and methylation analysis confirmed this. Methylation analysis of the intact LPS from each strain indicated absence of 3-substituted and 6-substituted D-Glc, 3-substituted DD-Hep and 6-substituted DD-Hep (for O:3::HP0479 and 26695::HP0479 LPS) and a significant decrease in 2-substituted DD-Hep, suggesting deficiencies in the core biosynthesis.

FAB-MS analysis in the positive mode of the permethylated LPS from each strain indicated the presence of primary glycosyl oxonium ions at m/z 260 $[GlcNAc]^+$

and m/z 434 $[Fuc,GlcNAc]^+$ and secondary glycosyl oxonium ions at m/z 228 (260-32) $[GlcNAc]^+$ and m/z 402 (434-32) $[Fuc,GlcNAc]^+$. This evidence together with the absence of the primary glycosyl oxonium ion at m/z 682 [Fuc, GlcNAc, Hep]$^+$ suggested that the mutant LPS structure was lacking DD-Hep residue which bridges O-chain and the core oligosaccharide in the respective parent LPS (Monteiro *et al*, *Eur. J. Biochem.* 267: 305-320 (2000), Logan *et al*, *Mol. Microbiol.* 35: 1168-1179 (2000)). LPS from SS1:: HP0479 and 26695 was delipidated and desalted following gel filtration chromatography on a Bio-Gel P-2 column. Fractions containing core oligosaccharide components were subjected to the mass spectrometric analysis using combined capillary zone electrophoresis-electrospray-mass spectrometry (CZE-ESMS) in the positive mode, followed by MS/MS analysis of the most abundant oligosaccharide fragments. The product ion spectrum showed two major singly charged fragment ions at m/z 1612 and m/z 1246, containing an anhydro-KDO. The fragment ion at m/z 1612 could be assigned to the glycoform $FucGlcNAcHex_2Hep_3(PE)KDO$ (Fig. 5), based on the linkage and FAB-MS analyses data and recent structural studies (Monteiro *et al*, *Eur. J. Biochem.* 267: 305-320 (2000)). The MS/MS spectrum of m/z 1246 was consistent with the core fragment $Hex_2Hep_3(PE)KDO$ as confirmed by a consecutive cleavage of glycosidic bonds yielding a direct sequence assignment. These structural assignments are consistent with the presence of 2,7-substituted DD-Hep, 7-substituted DD-Hep and 2-substituted DD-Hep in the methylation analysis of LPS mutants 26695::HP0479kan, SS1::HP0479kan, O:3::HP0479kan. Absence of the first DD-Hep which serves as a link between the O-chain and the core oligosaccharide and is glycosylated by 1,6-glucan, resulted in the loss of O-chain and DD-heptan (for serotype O:3). A comparison of the structures identified in parent and HP0479 mutant LPS is presented in Fig. 3.

Mouse Colonization Studies

The role of S-type LPS in gastric colonisation was investigated using the SS1 strain of *H. pylori* which others (Lee *et al*, *Gastroenterology* 112: 1386-1397 (1997); Ferrero *et al*, *Infect. Immun.* 66: 1349-1355 (1998); Conlan *et al*, *Can. J. Microbiol.* 45:975-980 (1999)) have shown to be capable of colonising the stomachs of mice, including the CD1 strain used in the present study. Both

parental SS1 and SS1 HP0826 mutant which was obtained by natural transformation w re used to orogastrically inoculate mice. The parent SS1 cells produce considerable amounts of S type LPS displaying Lewis Y epitopes while cells in which HP0826 has been inactivated produce a faster migrating, rough type LPS molecule no longer displaying Lewis epitopes. To minimise the likelihood that any observed differences in *in vivo* behaviour arose as a result of exogenous influences, care was taken to ensure that the mutant and parental strains underwent equivalent *in vitro* manipulations before being gavaged into mice. In an initial experiment, groups of mice were gavaged with either wild-type or mutated *H. pylori* SS1. Representative mice from each group were killed 6 or 12 weeks later and the stomach burdens of *H. pylori*, and level of *Helicobacter*-specific circulating immunoglobulin G determined. By 6 weeks of infection, 5.65 +/- 0.26 $\log_{10}$CFU (colony-forming units) of wild-type bacteria were recovered from the stomachs of mice (n=4) challenged with this organism, whereas only 4.27+/- 0.26 $\log_{10}$CFU of the mutant bacteria were recovered from the stomachs of mice gavaged with it. This 24-fold decreased recovery of mutant *versus* wild-type *H. pylori* SS1 was statistically significant according to the Mann-Whitney Rank Sum Test ($p<0.05$). Similarly, by 12 weeks there was a 10-fold difference in numbers of wild-type (5.81+/-0.51 $\log_{10}$CFU, n=5) and mutant (4.79+/-0.43 $\log_{10}$CFU, n=5) bacteria recovered, and this too was statistically significant ($p<0.05$). PCR performed on digested stomach tissue confirmed the above findings, indicating that the decreased recovery was not due to any innate unculturability of the mutant bacteria. Likewise, by 12 weeks of infection sera from mice infected with wild-type SS1 all reacted by ELISA against a sonicate of *H. pylori* as coating antigen (average IgG titre = 1270+/-2166) whereas only 3/5 mice infected with mutant SS1 had seroconverted (mean IgG titre of seropositives = 123+/-94). Additionally, when either parental or mutant LPS was used as the coating antigen in ELISA, only mice infected with the parental strain of H. pylori showed evidence of seroconversion.

To determine whether the colonisation differences observed in the aforementioned experiment were due to an initial inability of the mutant strain to colonise or due to its subsequent elimination, a complementary experiment

examined gastric colonization levels of parental and mutated *H. pylori* SS1 at 1 and 3 weeks post-challenge. By one-week post-challenge, 5.81+/-0.29 log_{10}CFU (n=5) of wild-type bacteria, but only 3.94+/-0.33 log_{10}CFU (n=5) of the mutant bacteria were recovered from the stomachs of the respectively infected mice. This 74-fold difference was statistically significant ($P< 0.05$) and convincingly shows that *H. pylori* SS1 bacteria unable to produce S-type LPS are significantly impaired in their ability to initially colonise the murine stomach. In this experiment, approximately 17-fold more wild-type than mutant *H. pylori* (5.4+/-0.34 log_{10} CFU, n=5 *versus* 4.18+/-0.14 log_{10}CFU, n=5) were recovered from the stomachs of relevant mice at three weeks of infection.

Results of mouse colonization experiments for the parent (SS1) strain of H. pylori and their mutant strains SS0826, SS0159 and SS0479 are summarized in Table 5.

Table 5. Mouse colonization data. Numbers in the table show levels of colonization of mice stomachs (as log_{10}CFU/stomach +/- standard deviation) after the indicated number of weeks (WK) of infection. ND: not determined BDL: less than 500 bacteria

	STRAIN	WK 1	WK 3	WK 6	WK 12	WK 20
EXP 1	SS1	5.81 +/- 0.29 (n = 5)	5.40 +/- 0.34 (n = 5)	5.65 +/- 0.26 (n = 4)	5.81 +/- 0.51 (n = 5)	ND
	SS0826	3.94 +/- 0.33 (n = 5)	4.18 +/- 0.17 (n = 5)	4.27 +/- 0.26 (n = 4)	4.79 +/- 0.43 (n = 5)	ND
EXP 2	SS1	5.43 +/- 0.03 (n = 4)	ND	ND	5.94 +/- 0.33 (n = 5)	5.84 +/- 1.10 (n = 5)
	SS0159	3.37 +/- 0.20 (n = 4)	ND	ND	3.09 +/- 0.42 (n = 5)	< 3.76 (n = 5)
EXP 3	SS1	4.76 +/- 0.93 (n = 5)	ND	ND	5.02 +/- 1.06 (n = 5)	ND
	SS0479	BDL (n = 5)	ND	ND	BDL (n=5)	ND

Exp 1: Individual mice inoculated by gavage on D1, D3, D6 with 0.2ml of broth grown cells suspended in PBS at cell concentration of ~1 x 10^{10}/ml.

Exp 2: Individual mice inoculated by gavage on D1 + D3 with 0.2ml of broth grown cells suspended in PBS at cell concentration of ~2 x 10^{10}/ml.

Exp 3: Individual mice inoculated by gavage on D1 and D3 with 0.2ml of broth grown cells suspended in PBS at cell concentration of $4.7x10^{10}$/ml (D1) and $1x10^{7}$/ml (D3)

The above data show that all the mutants with disrupted genes have a reduced ability to colonize the murine stomach, as compared with the parent strain. SS0479 strain (*H. pylori* strain SS1 having disrupted gene HP0479) is the least capable of colonization.

EXPERIMENTAL

Bacterial strains and culture conditions

Helicobacter pylori strain 26695 (Tomb *et al*, *supra*) used for the initial cloning was obtained from R. A. Alm, Astra, Boston. *H. pylori* strain SS1 was obtained from A. Lee. *H. pylori* reference strain ATCC43504 and *H. pylori* serogroup O:3 isolate were from J. Penner. PJ1 was a fresh clinical isolate of *H. pylori*. Helicobacter strains were grown on at 37°C on antibiotic supplemented (Lee *et al*, *supra*) trypticase soy agar plates containing 7% horse blood (GSS agar) in a microaerophilic environment for 48h (Kan 20 μg/ml). For growth in liquid culture, antibiotic supplemented Brucella broth containing 5% fetal bovine serum, was inoculated with *H. pylori* cells harvested from 48h trypticase soy agar/horse blood plates and incubated for 36h in a Trigas (Nuaire, Plymouth, MN) incubator (85% N_2, 10%CO_2, 5%O_2) on a shaking platform. *Escherichia coli* strain DH5α was used as host for plasmid cloning experiments and was grown on L-agar plates at 37°C supplemented with ampicillin (50μgml^{-1}) and/or kanamycin (20μgml^{-1})

β-1,4-galactosyltransferase activity

Glycosyltransferase assays were performed essentially as described previously (Gilbert *et al.*, *supra*). Cells were scraped from a 3 day old plate culture of *H. pylori*, the cells were stored frozen at -20°C. Cell extracts were made by mixing the cell pellet with 2 volumes of glass beads, and grinding with a ground glass pestle in the microcentrifuge tube. The paste was extracted twice with 50 µl of 50 mM MOPS-NaOH buffer pH 7.0. Reactions contained 0.5 mM FCHASE-aminophenyl-β-GlcNAc, 10 mM $MnCl_2$, 0.5 mM UDP-Gal, 50 mM MOPS-NaOH pH 7.0, and 10 µl of cell extract in a final volume of 20 µl. For reactions with the cell extracts of *H. pylori* the reactions were incubated 3-5 h at 37°C, whereas with the extracts containing the recombinant enzyme the reactions times were 30 – 60 min at 37°C. The TLC and CE analysis was performed as previously described (Gilbert *et al.*, *supra*). For TLC analysis 0.5 µl of the reaction mixture were spotted and developed and for CE analysis samples were diluted to an FCHASE-aminophenyl-β-GlcNAc concentration of 10 µM prior to analysis.

Recombinant DNA techniques and nucleotide sequence analysis

DNA sequencing of PCR products was performed using an Applied Biosystems (model 370A) automated DNA sequencer using the manufacturers cycle sequencing kit. All standard methods of DNA manipulation were performed according to the protocols of Sambrook *et al*, *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (1989). DNA probes for Southern blotting were labelled with DIG-11-dUTP using DIG-High Prime (Boehringer Mannheim, Montreal, Canada) and detection of hybridized probe with DIG Luminescent Detection Kit (Boehringer Mannheim Montreal, Canada). Primers used for the PCR gene amplification and for mutant constructs are shown in Table 6.

Table 6. Primer sequences for PCR amplification of HP0826, HP0159, HP0479 and HP1191 genes and for construction of respective mutant strains .

Primer	Primer (5'-3')sequence
HP0826-F1	cggatccGGTTTTTATAGCCATGATGC
HP0826-R1	cggatccAAGGCGGTTAAGTTTTGTTC
HP0826-mut1	TACAGATCGCTTCATTGAGTTCT
HP0826-mut2	CCAAGAGTTAGGCTATATCCGCTT
HP0159-F1	cgggatccTGTCAAATTCGCCTATAGCGT
HP0159-R1	cgggatccACCTATTTTAGGGAAACCGCT
HP0159-mut1	GCCGGGTTTTTAGTCGTGAAT
HP0159-mut2	AGGGAAAAGGCTTGACGAGG
HP0479-F1	GCCTTTATCAAGCTAGAG
HP0479-R1	CATAAATGTCCTAACAAGC
HP0479-mutF1	CAAAACCGCCAGGAGTTG
HP0479-mutR1	GGTTATGGGAATGAATTTGG
HP1191-F1	cgggatccCGGTCTTTAAACCCGCTCAACA
HP1191-R1	cgggatccCCGCTCTTCTCACGCCTTTAA

Site specific mutagenesis of HP0826

HP0826 clone of *Helicobacter pylori* strain 26695 was mutagenized in *E. coli* by ligation of the Km^r cassette described by Labigne *et al* (*J. Bacteriol.* 170: 1704-1708 (1988)) to pUC19 containing the HP0826 gene. Deletion of a central 66bp region of the gene was achieved by reverse PCR (Pwo polymerase, Boehringer Mannheim) using the outward primers 5'TACAGATCGCTTCATTGAGTTCT3' and 5'CCAAGAGTTAGGCTATATCCGCTT3' followed by blunt end ligation with the Km^r cassette. The mutated allele was returned to Helicobacter by natural transformation according to the method of Haas *et al* (*supra*).

Electrophoresis and Western blotting

SDS-PAGE was performed with a mini-slab gel apparatus (Biorad) by the method of Laemmli (*Nature* 227: 680-685 (1970)). LPS samples were prepared from whole cells according to a previously described method (Logan *et al*, *Infect. Immun.* 45: 210-216 (1984)), equivalent amounts loaded in each lane and stained according to Tsai *et al* (*Anal. Biochem.* 119: 115-119 (1982)) or transferred to nitrocellulose for immunological detection as previously described

(Logan *et al, supra*). Anti Lewis monoclonal antibodies (Signet Laboratories Inc, Dedham, MA) were used at 1:500 dilution.

Isolation of membrane fraction

Broth grown cells (18h) were harvested and resuspended in 20mM Tris (pH 7.4). Following sonication (3x60sec) intact cells were removed by centrifugation at 4000xg, and membranes sedimented by centrifugation at 40,000xg, washed in 20mM Tris (pH7.4) recentrifuged, and resuspended in 0.5ml 20mM Tris (pH7.4). Equivalent amounts of SS1, 26695 parent and mutant strains were analyzed by SDS-PAGE and stained by Coomassie Blue.

Isolation of Lipopolysaccharides

The LPSs were isolated by the hot phenol-water extraction procedure (Westphal *et al, Meth. Carbohydr. Chem.* 5: 83-91 (1965)). The LPSs were purified by gel-permeation-chromatography on a column of Bio-Gel P-2 (1m x 1cm) with water as eluent. In all cases, only one carbohydrate positive fraction was obtained which eluted in the high M_r range (Dubois *et al, Anal. Chem.* 28: 350-356 (1956)). These intact *H. pylori* LPSs then were used for chemical analyses.

Sugar Composition and Methylation Linkage Analyses

Sugar composition analysis was performed by the alditol acetate method (Sawardeker *et al, Anal. Chem.* 39:1602-1604 (1967)). The hydrolysis was done in 4M trifluoroacetic acid at 100°C for 4h or 2M trifluoroacetic acid at 100°C for 16h followed by reduction in H_2O with $NaBD_4$, and subsequent acetylation with acetic anhydride and with residual sodium acetate as the catalyst. Alditol acetate derivatives were analyzed by gas-liquid-chromatography mass-spectrometry (GLC-MS) using a Hewlett-Packard chromatograph equipped with a 30 m DB-17 capillary column [210°C (30 min) to 240°C at 2°C/min] and MS in the electron impact (EI) mode was recorded using a Varian Saturn II mass spectrometer. Methylation linkage analysis was carried out by the NaOH/DMSO/CH_3I procedure (Ciucanu *et al, Carbohydr. Res.* 131: 209-217 (1984)) and with characterization of permethylated alditol acetate derivatives by GLC-MS in the EI mode (DB-17 column, isothermally at 190°C for 60 min).

Fast Atom Bombardment-Mass Spectrometry (FAB-MS)

A fraction of the methylated sample was used for positive ion fast atom bombardment-mass spectrometry (FAB-MS) which was performed on a Jeol JMS-AX505H mass spectrometer with glycerol(1) : thioglycerol(3) as the matrix. A 6 kV Xenon beam was used to produce pseudo molecular ions which were then accelerated to 3kV and their mass analyzed. Product ion scan (B/E) and precursor ion scan (B^2/E) were preformed on metastable ions created in the first free field with a source pressure of $5x10^{-5}$ torr. The interpretations of positive ion mass spectra of the permethylated LPS derivatives were as previously described by Dell *et al* (Carbohydr. Res. 200: 59-67 (1990).

Electrospray mass spectrometry

Samples were analyzed on a crystal Model 310 CE instrument (ATI Unicam, Boston, MA, USA) coupled to an API 3000 mass spectrometer (Perkin-Elmer/Sciex, Concord, Canada) via a microIonspray interface. A sheath solution (isopropanol-methanol, 2:1) was delivered at a flow rate of 1 μL/min to a low dead volume tee (250 μm i.d., Chromatographic Specialties, Brockville, Canada). All aqueous solutions were filtered through a 0.45-μm filter (Millipore, Bedford, MA, USA) before use. An electrospray stainless steel needle (27 gauge) was butted against the low dead volume tee and enabled the delivery of the sheath solution to the end of the capillary column. The separation were obtained on about 90 cm length bare fused-silica capillary using 10 mM ammonium acetate/ammonium hydroxide in deionized waster, pH 9.0, containing 5% methanol. A voltage of 25 kV was typically applied at the injection. The outlet of the capillary was tapered to ca. 15 μm i.d. using a laser puller (Sutter Instruments, Novato, CA, USA). Mass spectra were acquired with dwell times of 3.0 ms per step of 1 *m/z* unit in full-mass-scan mode. For CZE-ES-MS/MS experiments, about 30 nL sample was introduced using 300 mbar for 0.1 min. The MS/MS data were acquired with dwell times of 1.0 ms per step of 1 *m/z* unit. Fragment ions formed by collision activation of selected precursor ions with nitrogen in the RF-only quadrupole collision cell, were mass-analyzed by scanning the third quadrupole. Collision energies were typically 60 eV (laboratory frame of reference).

M use Colonization

Specific Pathogen free Female CD1 mice were purchased from Charles Rivers Laboratories, Montreal when they were 6-8 weeks old. Mice were maintained and used in accordance with the recommendations of the Canadian Council on Animal Care, Guide to the Care and Use of Experimental Animals (1993). Mice were inoculated with bacteria harvested from 36h broth culture. Aliquots of 0.2 ml, containing approximately 10^8 bacteria resuspended in PBS were given by gavage directly into the gastric lumen using a 20g gavage needle. Three inocula were given over a period of 6 days. No attempt was made to neutralize gastric acidity prior to inoculation. To recover viable bacteria from the stomach, mice were killed by CO_2 asphyxiation, and their stomachs removed whole. Stomachs were cut open along the greater curvature, and the exposed lumenal surface was gently irrigated with 10 ml of sterile PBS, delivered via a syringe fitted with a 20g gavage needle, to dislodge the loosely adherent stomach contents. This step effectively diminished the small numbers of ubiquitous contaminating bacteria that otherwise overgrow on GSS agar to thereby mask the presence of the slower growing *H. pylori* organisms. The washed stomach tissue was then homogenised, and serial dilutions plated on GSS agar. *H. pylori* colonies were counted following 3-6 days incubation.

Detection of *H. pylori* specific antibodies by ELISA

Sera for antibody determinations were prepared from clotted blood obtained from a lateral tail vein during the course of an experiment or by cardiac puncture at the time of necropsy. Sera were screened for the presence of specific IgG isotype anti- *H. pylori* antibodies by ELISA essentially by the method of Engvall *et al* (*J. Immunol.* 109: 129-135 (1972)). Briefly, microtitre plates (Dynatech Immunolon II) were coated with 100 µl antigen (50 µg protein/ml in 0.05M carbonate buffer pH 9.8) and incubated overnight at 4°C. Antigen was prepared by resuspending plate grown *H. pylori* in PBS and sonicating the suspension until a translucent solution was obtained. The sonicate was membrane filter sterilized through a 0.45 µm filter. The protein content of the filtrate was determined by Lowry assay using a commercial kit. Sodium azide was added to 0.05% w/v and the antigen

solution was stored at 4°C. When LPS was used as the coating antigen th concentration was 10μg/ml. Sera were screened at a starting dilution of 1/40 and were titrated through a two-fold dilution series down a column of 8 wells. The developing antibody was goat-anti-mouse IgG conjugated to alkaline phosphatase (Caltag Laboratories). Titres were determined from plots of absorbance at 410 nm *versus* dilution and were defined as the reciprocal of the dilution giving an A_{410} equivalent to 0.25. Standard negative and positive control sera identified by a preliminary ELISA of candidate samples were included on each plate. Titres were analysed statistically by Mann Whitney Rank Sum Test and were considered to be significantly different to comparative samples when p values <0.05 were obtained.

Although various particular embodiments of the present invention have been described hereinbefore for purposes of illustration, it would be apparent to those skilled in the art that numerous variations may be made thereto without departing from the spirit and scope of the invention, as defined in the appended claims.

WHAT IS CLAIMED IS:

1. An isolated or recombinant polynucleotide encoding at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

2. A polynucleotide according to claim 1, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the LPS.

3. A polynucleotide according to claim 1, wherein the glycosyltransferase is involved in the biosynthesis of the core region of the LPS.

4. A polynucleotide according to claim 2, wherein the glycosyltransferase is a galactosyltransferase.

5. A polynucleotide according to claim 4, wherein the galactosyltransferase is a β-1,4-galactosyltransferase.

6. A polynucleotide according to claim 5, wherein the *Helicobacter* is a strain of *H. pylori*.

7. A polynucleotide according to claim 3, wherein the glycosyltransferase is a glucosyltransferase.

8. A polynucleotide according to claim 7, wherein the glycosyltransferase is an α-1,6-glucosyltransferase.

9. A polynucleotide according to claim 8, wherein the *Helicobacter* is a strain of *H. pylori*.

10. A polynucleotide according to claim 3, wherein the glycosyltransferase is a heptosyltranferase.

11. A polynucleotide according to claim 10, wherein the heptosyltransferase is an ADP-heptose-LPS heptosyltransferase II.

12. A polynucleotide according to claim 11, wherein the *Helicobacter* is a strain of *H. pylori*.

13. An isolated or recombinant polynucleotide having sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 and fragments and variants thereof.

14. An isolated or recombinant polynucleotide having at least about 70% identity to the polynucleotide according to claim 13.

15. An isolated or recombinantly produced polypeptide comprising at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

16. A polypeptide according to claim 15, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the LPS.

17. A polypeptide according to claim 15, wherein the glucosyltransferase is involved in the biosynthesis of the core region of the LPS.

18. A polypeptide according to claim 16, wherein the glycosyltransferase is a galactosyltransferase.

19. A polypeptide according to claim 18, wherein the galactosyltransferase is a β-1,4-galactosyltransferase.

20. A polypeptide according to claim 19, wherein the *Helicobacter* is a strain of *H. pylori*.

21. A polypeptide according to claim 17, wherein the glycosyltransferase is a glucosyltransferase.

22. A polypeptide according to claim 21, wherein the glucosyltransferase is an α-1,6-glucosyltransferase.

23. A polypeptide according to claim 22, wherein the *Helicobacter* is a strain of *H. pylori*.

24. A polypeptide according to claim 17, wherein the glycosyltransferase is a heptosyltranferase.

25. A polypeptide according to claim 24, wherein the heptosyltransferase is an ADP-heptose-LPS heptosyltransferase II.

26. A polypeptide according to claim 24, wherein the *Helicobacter* is a strain of *H. pylori*.

27. An isolated or recombinantly produced polypeptide having sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16 and fragments and variants thereof.

28. An isolated or recombinantly produced polypeptide having at least about 50% identity to the isolated polypeptide according to claim 27.

29. A recombinant vector comprising a nucleic acid encoding at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

30. A recombinant vector according to claim 29, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the LPS.

31. A recombinant vector according to claim 29, wherein the glycosyltransferase is involved in the biosynthesis of the core region of the LPS.

32. A recombinant vector according to claim 30, wherein the glycosyltransferase is a galactosyltransferase.

33. A recombinant vector according to claim 32, wherein the galactosyltransferase is a β-1,4-galactosyltransferase.

34. A recombinant vector according to claim 33, wherein the *Helicobacter* is a strain of *H. pylori*.

35. A recombinant vector according to claim 31, wherein the glycosyltransferase is a glucosyltransferase.

36. A recombinant vector according to claim 35, wherein the glucosyltransferase is an α-1,6-glucosyltransferase.

37. A recombinant vector according to claim 36, wherein the *Helicobacter* is a strain of *H. pylori*.

38. A recombinant vector according to claim 31, wherein the glycosyltransferase is a heptosyltranferase.

39. A recombinant vector according to claim 38, wherein the heptosyltransferase is an ADP-heptose-LPS heptosyltransferase II.

40. A recombinant vector according to claim 39, wherein the *Helicobacter* is a strain of *H. pylori*.

41. A recombinant vector according to claim 29, wherein the glycosyltransferase has a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16 and fragments and variants thereof.

42. An expression cassette that comprises a nucleic acid encoding at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

43. An expression cassette according to claim 42, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the LPS.

44. An expression cassette according to claim 42, wherein the glycosyltransferase is involved in the biosynthesis of the core region of the LPS.

45. An expression cassette according to claim 42, wherein the glycosyltransferase has a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16 and fragments and variants thereof.

46. A host cell comprising a recombinant nucleic acid which can express a protein encoding at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

47. A host cell according to claim 46, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the LPS.

48. A host cell according to claim 46, wherein the glycosyltransferase is involved in the biosynthesis of the core region of the LPS.

49. A host cell according to claim 46, wherein the glycosyltransferase has a sequence selected from th group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 and fragments and variants thereof.

50. A host cell according to claim 49, wherein the cell is a eukaryotic cell.

51. A host cell according to claim 49, wherein the cell is a prokaryotic cell.

52. A host cell according to claim 51, wherein the prokaryotic cell is a cell of *E. coli*.

53. A method for producing a polypeptide comprising at least a portion of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS), comprising the steps of maintaining a host cell of claim 46 under conditions suitable for expression of said polypeptide and recovering the polypeptide so produced.

54. A method according to claim 53, wherein the glycosyltransferase has a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 and fragments and variants thereof.

55. A method according to claim 53, further including the step of purifying the recovered polypeptide.

56. A hybridization probe comprising a portion of a polynucleotide encoding a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

57. A hybridization probe according to claim 56, wherein the glycosyltransferase has a sequence selected from the group consisting of

SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15 and fragments and variants thereof.

58. A hybridization probe according to claim 57, wherein the probe comprises at least about 15 nucleotides.

59. A mutant strain of *H. pylori*, said mutant strain having deactivated at least one gene encoding a glycosyltransferase involved in the biosynthesis of a *H. pylori* lipopolysaccharide (LPS).

60. A mutant strain according to claim 59, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain of the LPS.

61. A mutant strain according to claim 59, wherein the glycosyltransferase is involved in the biosynthesis of the core region of LPS.

62. A mutant according to claim 59, wherein the glycosyltransferase is coded by open reading frames 0826, 0159, 0479 or 1191.

63. A vaccine composition comprising an antigen derived from a mutant strain of *H. pylori* according to claim 59.

64. A vaccine composition according to claim 63, wherein the antigen is an at least partially purified lipopolysaccharide.

65. A vaccine composition according to claim 64, wherein the antigen is conjugated to a protein.

66. A live attenuated vaccine composition comprising a mutant strain of *H. pylori* according to claim 59.

67. A reaction mixture for an enzymatic synthesis of a *Helicobacter* lipopolysaccharide or a portion thereof, th mixture comprising an isolated polypeptide having activity of a *Helicobacter* glycosyltransferase involved in the biosynthesis of a *Helicobacter* lipopolysaccharide (LPS).

68. A reaction mixture according to claim 67, wherein the glycosyltransferase is involved in the biosynthesis of the O-chain region of the *Helicobacter* lipopolysaccharide.

69. A reaction mixture according to claim 67, wherein the glycosyltransferase is involved in the biosynthesis of the core region of the *Helicobacter* lipopolysaccharide.

70. A reaction mixture according to claim 66, wherein the bacterial lipopolysaccharide is a mimic of a *Helicobacter* lipopolysaccharide.

1/5

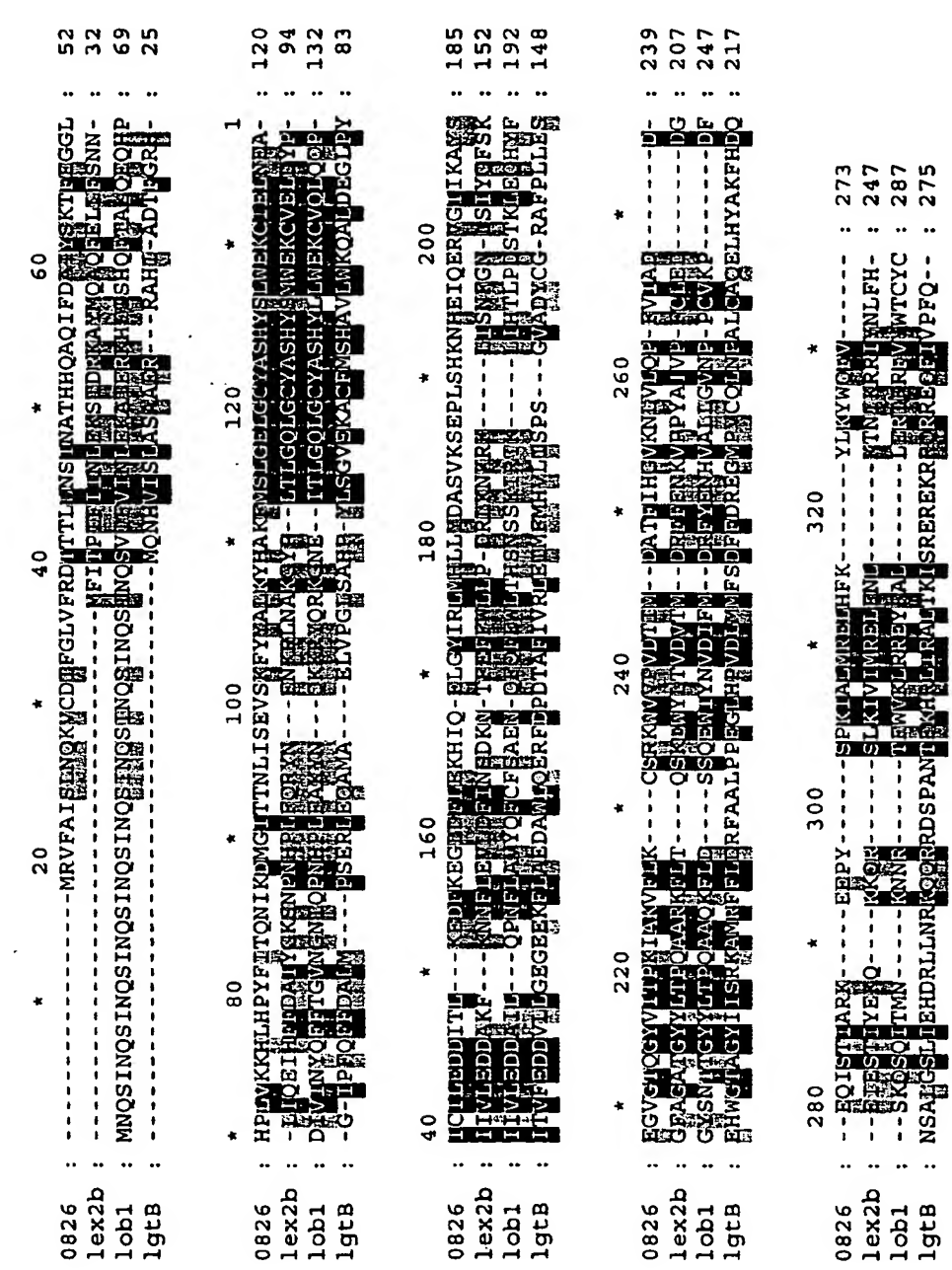


Fig. 1

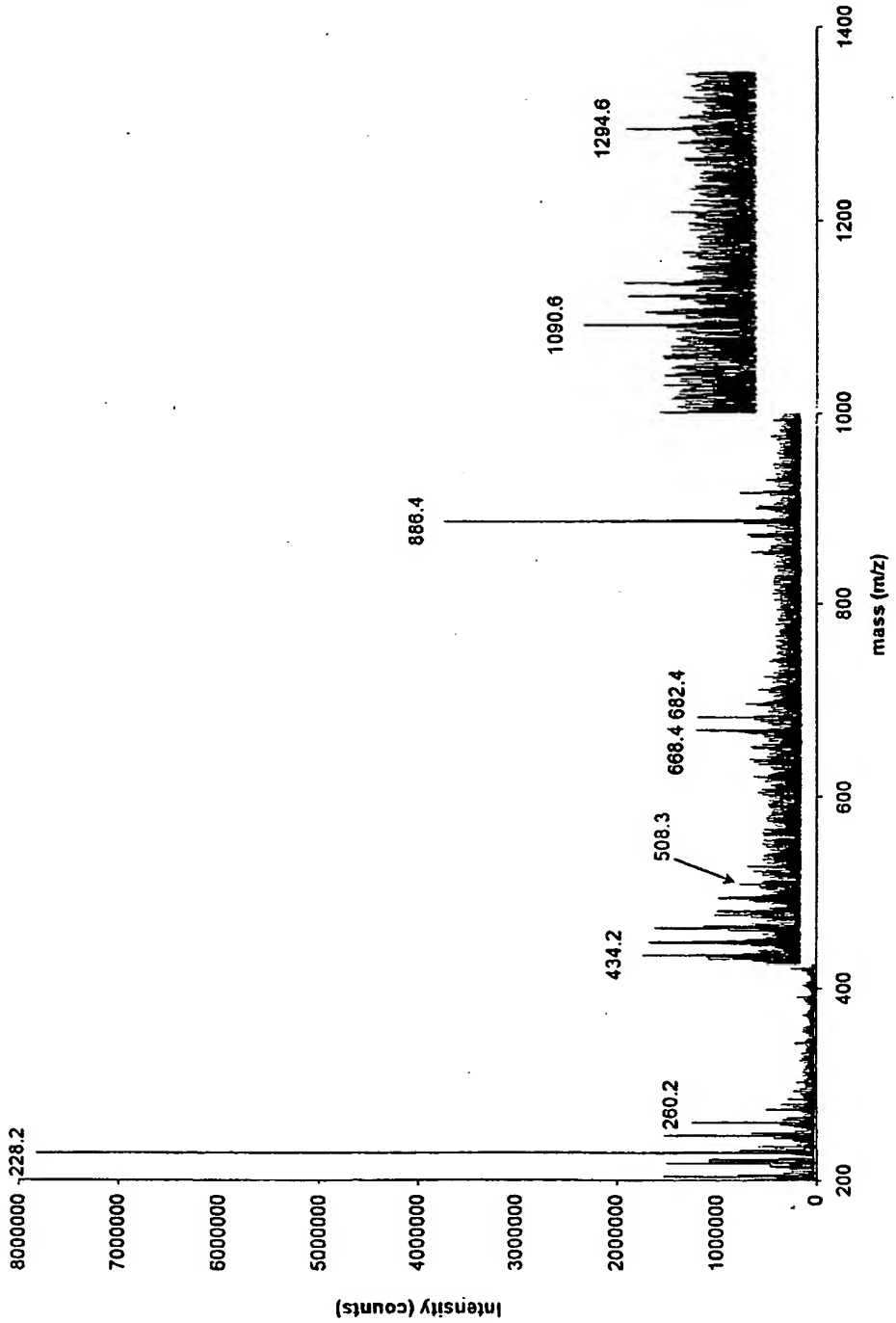
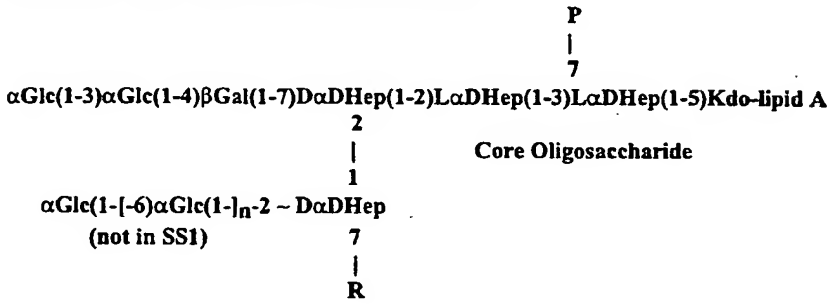

Intensity (counts)
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
mass (m/z)
200
400
600
800
1000
1200
1400
228.2
260.2
434.2
508.3
668.4
682.4
886.4
1090.6
1294.6

Fig. 2

***H. pylori* 26695 and SS1 LPS structure.**

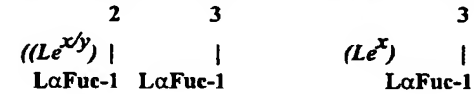


O-chain regions

R= βGal(1-4)βGlcNAc(1-[-3)βGal(1-4)βGlcNAc(1-]n-{-3)DDHep(1-, -2)DDHep(1-, -6)DDHep(1-}-

Strain 26695

R= βGal(1-4)βGlcNAc(1-[-3)βGal(1-4)βGlcNAc(1-]n-



Strain SS1

***H. pylori* HP0826 mutant lipopolysaccharides of strains 26695, SS1**

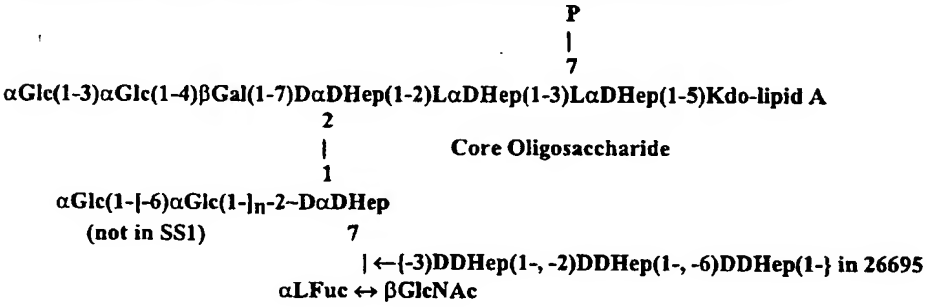
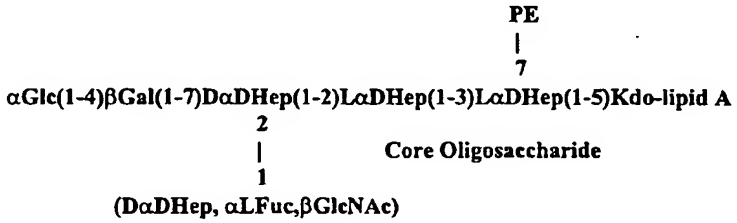


Fig. 3

***H. pylori* HP0159 mutant lipopolysaccharides of strains 26695 and SS1 .**



***H. pylori* HP0479 mutant lipopolysaccharides of strains 26695 and SS1.**

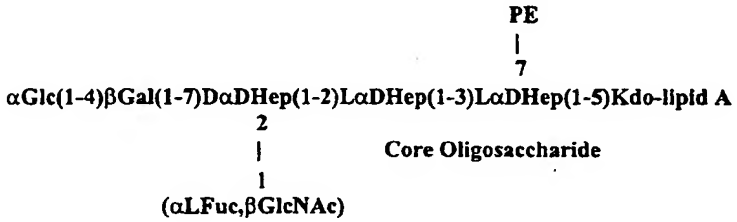


Fig. 3 (Cont.)

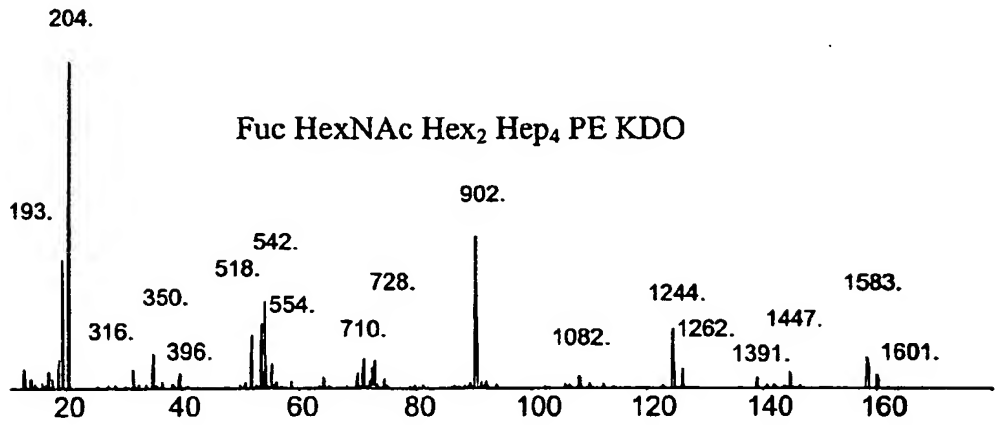

Fuc HexNAc Hex2 Hep4 PE KDO
204.
193.
316.
350.
396.
518.
542.
554.
710.
728.
902.
1082.
1244.
1262.
1391.
1447.
1583.
1601.
20
40
60
80
100
120
140
160

Fig. 4

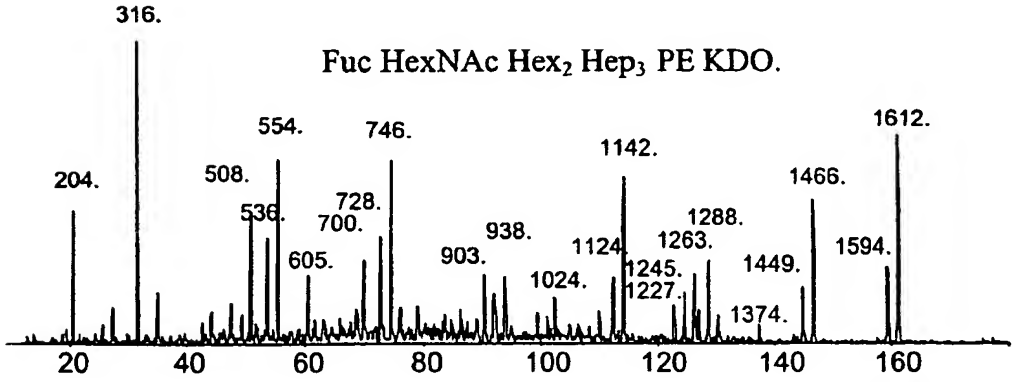

Fuc HexNAc Hex2 Hep3 PE KDO.
316.
204.
508.
536.
554.
605.
700.
728.
746.
903.
938.
1024.
1124.
1142.
1227.
1245.
1263.
1288.
1374.
1449.
1466.
1594.
1612.
20
40
60
80
100
120
140
160

Fig. 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00796 A3

(51) International Patent Classification[7]: C12N 9/10, 1/21, C12Q 1/68, C12P 19/44, A61K 39/106, 31/739 // (C12N 9/10, C12R 1:01) (C12N 1/21, C12R 1:19, 1:01)

(21) International Application Number: PCT/CA00/00777

(22) International Filing Date: 28 June 2000 (28.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/140,820 28 June 1999 (28.06.1999) US

(71) Applicant *(for all designated States except US)*: NATIONAL RESEARCH COUNCIL OF CANADA [CA/CA]; Intellectual Property Services, Building M-58, Room EG-12, Montreal Road, Ottawa, Ontario K1A 0R6 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LOGAN, Susan, M. [CA/CA]; 62 Kilbarry Crescent, Ottawa, Ontario K1K 0H1 (CA). WAKARCHUK, Warren [CA/CA]; 11-837 Eastvale Drive, Gloucester, Ontario K1J 7T5 (CA). CONLAN, Wayne [CA/CA]; 1152 Bordeau Grove, Orleans, Ontario K1C 2M7 (CA). MONTEIRO, Mario, A. [CA/CA]; 171 O'Connor Street, Apt. 105, Ottawa, Ontario K1P 2T4 (CA). ALTMAN, Eleonora [CA/CA]; 689 Laverendrye Drive, Gloucester, Ontario K1J 7X5 (CA). HIRATSUKA, Koji [CA/CA]; 11 Hedgerow Lane, Stittsville, Ontario K2S 1C9 (CA).

(74) Agent: ANDERSON, J., Wayne; National Research Council of Canada, Intellectual Property Services, Building M-58, Room EG12, Montreal Road, Ottawa, Ontario K1A 0R6 (CA).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
9 August 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00796 A3

(54) Title: GLYCOSYLTRANSFERASES OF *HELICOBACTER* PYLORI

(57) Abstract: Novel isolated polynucleotides encoding glycosyltransferases involved in the biosynthesis of the lipopolysaccharide of *Helicobacter pylori*, together with recombinant DNA constructs and vectors containing polynucleotide sequences encoding such glycosyltransferases are disclosed. These nucleic acid constructs and vectors may be used for the preparation of glycosyltransferases they encode, by expressing the coding polynucleotide sequences in suitable host cells. Also disclosed are isolated polypeptides having enzymatic activity of helicobacterial glycosyltransferases. Such polypeptides are particularly useful for screening of modulators of their enzymatic activity, in particular enzymatic inhibitors having potential antibacterial activity.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA [redacted]00777

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N9/10 C12N1/21 C12Q1/68 C12P19/44 A61K39/106
A61K31/739 //(C12N9/10,C12R1:01),(C12N1/21,
C12R1:19,1:01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q C12P A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, STRAND, BIOSIS, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TOMB J -F ET AL: "THE COMPLETE GENOME SEQUENCE OF THE GASTRIC PATHOGEN HELICOBACTER PYLORI" NATURE,GB,MACMILLAN JOURNALS LTD. LONDON, vol. 388, no. 6642, 7 August 1997 (1997-08-07), pages 539-547,TABEL, XP002062106 ISSN: 0028-0836 cited in the application table 2 -& DATABASE EMBL [Online] Accession AE000594, 25 August 1997 (1997-08-25) TOMB J -F ET AL: "Helicobacter pylori 26695 section 72 of 134 of the complete genome." XP002155934 100% identity in full length overlap with SEQ ID NO 1 (Pos. 1551-2372); 99.6% -/--	1,13-15, 27-29, 41,42, 45,46, 49-58

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 December 2000	21.03.01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lejeune, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	identity in full length overlap between translated amino acid sequence and SEQ ID NO 2 96.3% identity in 821 BP overlap with SEQ ID NO 9 (Pos. 1551-2371); 94.5% identity in 273 AA overlap between translated amino acid sequence and SEQ ID NO 2 ---	
P,X	WO 99 40205 A (ENDO TETSUO ;KOIZUMI SATOSHI (JP); OZAKI AKIO (JP); TABATA KAZUHIK) 12 August 1999 (1999-08-12) 96.5% identity in 819 BP overlap between SEQ ID NO 1 and SEQ ID NO 2 of EP1054062; 96.7% identity in full length overlap between the corresponding amino acid sequences; 97.2% identity in 819 BP overlap between SEQ ID NO 9 and SEQ ID NO 2 of EP1054062; 96.7 % identity between the corresponding amino acid sequences. -& EP 1 054 062 A (KYOWA HAKKO KOGYO KK) 22 November 2000 (2000-11-22) column 1, paragraph 3 abstract ---	1,2,4-6, 13-16, 18-20, 27-30, 32-34, 41-43, 45-47, 49-58, 67,68
X	WANG G ET AL: "MOLECULAR GENETIC BASIS FOR THE VARIABLE EXPRESSION OF LEWIS Y ANTIGEN IN HELICOBACTER PYLORI: ANALYSIS OF THE ALPHA(1,2) FUCOSYLTRANSFERASE GENE" MOLECULAR MICROBIOLOGY,GB,BLACKWELL SCIENTIFIC, OXFORD, vol. 31, no. 4, February 1999 (1999-02), pages 1265-1274, XP000889904 ISSN: 0950-382X the whole document ---	1,2,15, 16,29, 30,42, 43,46, 47,53, 55,56, 59,60, 67,68
X	MARTIN S L ET AL: "Lewis X biosynthesis in Helicobacter pylori" JOURNAL OF BIOLOGICAL CHEMISTRY,US,AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, vol. 272, no. 34, 22 August 1997 (1997-08-22), pages 21349-21356, XP002085414 ISSN: 0021-9258 the whole document --- -/--	1,2,15, 16,29, 30,42, 43,46, 47,53, 55,56, 67,68

INTERNATIONAL SEARCH REPORT

International Application No
PC[illegible]/CA[illegible]/00777

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHAN N W ET AL: "THE BIOSYNTHESIS OF LEWIS X IN HELICOBACTER PYLORI" GLYCOBIOLOGY,GB,IRL PRESS,, vol. 5, no. 7, 1995, pages 683-688, XP002920175 ISSN: 0959-6658 cited in the application the whole document	67,68
X	WO 96 40893 A (ASTRA AB ;BERGLINDH O THOMAS (SE); MELLGAERD BJOERN L (SE); SMITH) 19 December 1996 (1996-12-19) abstract page 39 94.7 % identity in 819 BP overlap between SEQ ID NO 1 and SEQ ID NO 1436 of WO9640893; 94.8% identity in 273 AA overlap between SEQ ID NO 2 and SEQ ID NO 1887 of WO9640893 94.3 % identity in 819 BP overlap between SEQ ID NO 9 and SEQ ID NO 1436 of WO9640893; 93.8% identity in 273 AA overlap between SEQ ID NO 2 and SEQ ID NO 1887 of WO9640893	13,27,28

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int[redacted]ation No.
P[redacted]CA 00/00777

Box I Obs rvati ns where certain claims w re found unsearchable (Continuati n of it m 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Please see additional sheet, Invention 1.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210.

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 2,4-6,16,18-20,30,32-34,43,47,60,
68 (all completely) 1,13-15,27-29,41,42,45,46,
49-59,62-67,70 (all partially)

An isolated or recombinant polynucleotide encoding at least a portion of a Helicobacter galactosyltransferase involved in the biosynthesis of a Helicobacter lipopolysaccharide. Also, a polypeptide encoded by the polynucleotide, a vector comprising the polynucleotide, a host cell comprising the polynucleotide, a mutant strain having deactivated said galactosyltransferase, a vaccine derived from such a mutant and a reaction mixture with said galactosyltransferase. A polynucleotide having SEQ ID NO 1 or 9.

2. Claims: 7-9,21-23,35-37 (all completely) 1,3,13-15,17,
27-29,31,41,42,44-46,48-59,61-67,69,
70 (all partially)

Same as invention 1, but pertaining to glucosyltransferase and SEQ ID NO 3 or 11.

3. Claims: 10-12,24-26,38-40 (all completely) 1,3,13-15,17,
27-29,31,41,42,44-46,48-59,61-67,69,
70 (all partially)

Same as invention 1, but pertaining to heptosyltransferase and SEQ ID NO 5, 7, 13 or 15.

INTERNATIONAL SEARCH REPORT

ormation [redacted] nt family members

International [redacted]lication No
PCT/CA [redacted]/00777

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9940205 A	12-08-1999	JP 11221079 A	17-08-1999
		AU 2075599 A	23-08-1999
		EP 1054062 A	22-11-2000
WO 9640893 A	19-12-1996	AU 710880 B	30-09-1999
		AU 6327896 A	30-12-1996
		BR 9609430 A	24-08-1999
		CA 2223395 A	19-12-1996
		CN 1186516 A	01-07-1998
		CZ 9703886 A	12-08-1998
		EP 0842270 A	20-05-1998
		HU 9900766 A	28-06-1999
		JP 11504220 T	20-04-1999
		NO 975745 A	09-02-1998
		PL 324826 A	22-06-1998
		SK 165197 A	11-01-1999
		TR 9701538 T	21-03-1998
		AU 1055497 A	11-06-1997
		WO 9719098 A	29-05-1997